

|CURIEUX|

ACADEMIC JOURNAL

November Issue

Part 1 Issue 44

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: The Proportional Relationship Between Topographical Elevation Levels and Eutrophication-Induced Harmful Algal Blooms By Brianna Estocapio

Page 17: Duolingo Hindi and Short Term Memory Retention By Sai Smriti Saiganesh

Page 29: The Impact of Artificial Intelligence on the Landscape of Law By Nicole Pelletier

Page 40: Using Machine Learning to Bet on the Spread of NFL Games By Pranav Aiyar

Page 51: The Crucial Role of Eicosanoids in Liver Disease By Amber Wang

Page 56: Japanese Imperialism Through the Lens of Post-Colonial Theory By Diying (Eva) Jiang

Page 71: A 3-tiered Color Scale Nitrate Biosensor By Dongmin Choi

Page 86: Tourism Dynamics and Economic Growth in Norway: A Dual-Perspective Survey Analysis By Maanav Chittireddy

Page 96: The Familiar Melody of Memorization: A Musical Approach to Emily Dickinson By Tessa Gumberg

Page 100: Liberalism Wins Most Hated Doctrine of the Year By Tessa Gumberg

Page 104: On Morality: Is Morality Purely a Construct of Cultural Tradition? By Daniel Yu

Page 116: Fine-Tuning Neural Networks for Enhanced Emotion Detection: Insights from CNN, ResNet-50, DenseNet-121, and ViT Models By Amogh Singh

Page 132: Is peace in the West Bank and the Gaza Strip possible? By Ganning Ye

Page 137: Exploring Motivational Factors in High School Computer Science: A Quantitative Survey Study of Students' Expertise Levels of Java Programming in a Williamson County High School By Krithik Sama

Page 150: The Efficacy of AI Voice Agents in Enhancing Mental Health Outcomes: A Preliminary Investigation By Dhruv Reddy, Rithvik Rathinsabapathy

Page 158: Establish Data Protection To Ensure Privacy In the Surveillance Era By Audrey Horn

Page 165: Overview of Cerebrospinal Fluid, Glymphatic System, and Diagnosis of Multiple Sclerosis Through Machine Learning Extrapolation By Ethan Ngo

Page 183: Doping Agents in Sports By ZhongYu Richard Xun

Page 201: The Nanjing Massacre's Effect on Contemporary Chinese Views of Japan By Jiayi Li

Page 207: Making of an Athlete: Tissue-Specific Analysis of Genetic Regulators of Athletic Performance By Sri Riddhi Guruzu

Page 238: Decoding Fragile X Syndrome: Genetic and Epigenetic Modifications of FMR1 Gene and Therapeutic Implications By Mai Lam

Page 251: A Comparative Analysis of the Effect of Liquidity on the Price of Bitcoin By Rushil Jaiswal

Page 265: Estimating the Speed and Decay Rate of Muons By Aritro Ganguly

Page 274: Immunization in the US Army during World War II By Bella Gupta

Page 283: Simulation of Amyloid Beta Peptide Vaccine to Treat Alzheimer's By Yasmina Kodeih

Page 291: The Evolution, Impacts, and Future Prospects of Genetically Modified Organisms By Xiaoyu Zou

Page 305: New Frontiers for Sustainability in the Automotive Industry: the Case of Fuel Cells By Fahsai Pibulsonggram

Page 319: In which Ways do Superblocks Reflect on and Integrate with the Fundamental Values of the Local Mediterranean/Spanish cultures? By Lucas Srba Cabeza

Page 331: Can Super High School Construction Promote Economic Growth? Panel Data Based on Provincial Level By Xiaobo Zhou

Page 342: Estimating Bulgaria's 2023 Bright Economy Composite Index: Insights Into the Size of Hidden Economy By Aleksandar Georgiev

Page 363: A Systematic Review of Past, Current, and Future Approaches to Treat Breast Cancer By Frances Pappas

Page 376: The Origins of Multicellularity: Examining the Genetic and Environmental Factors Behind Multicellularity By Nakul Narayanan

Page 387: Why Do Civilisations Collapse? Is Our Civilisation in Danger? By Yanjia Yang

Page 394: Teenage Addiction: How Social Media and the "Doom Scroll" has Contributed to Mental Ailment By Jesleen Walia

Page 401: The Impact of AI on the Nursing Labor Market By Bill Wang

Page 424: Implementation of Grover's Search Algorithm in Quality of Service Path Selection for the Border Gateway Protocol By Shashank Garag

Page 441: The Relationship Between Video Game Use and Attention-Hyperactivity Disorder (ADHD) Among Children and Adolescents By Oruebubechukwu Asogwa

Page 459: Justification of the Decision to Begin the First Crusade By Merrick Jiang

Page 463: Chronic Kidney Disease in the Middle East By Michael Fam

Page 477: Investigating Carbon Markets: Mechanisms, Economics, and Implementation Strategies for a Sustainable Future By Timothy Liang

Page 492: Simulating the Effects of Weighted Dead Hangs on Spacing and Stretch of Glenohumeral Joints By Ian Lee

Page 501: A Novel Approach to Treat Parkinson's Disease By Emily Zhang

Page 506: Isolation of S. Mutans Bacteriophage and Investigation of S. Mutans Metabolic Versatility Across Media By Ang L. Li

Page 515: What Is the Intent of the New Mindreading Neurotech, And How Would It be Used? By Jahnvi Ponnolu

Page 521: Is Silliness Required for an Economy? By Vihan Agrawal

**Page 536: The Impact of Biological and Societal Factors on Individuals with ADHD
By Ava Bowles**

Page 552: Comparing Precision of Convolutional Neural Networks (CNN) Model on Audio Deepfake Detection Across English and Russian Languages By Saahil Manekar

Page 564: Using Prudence to Predict the International Electric Vehicle Market By Ryon Das

Page 580: Expanding Affordable Housing in Northern Virginia: Addressing Gaps in “Live, Work, Play” Communities By Sophie Zhang

The Proportional Relationship Between Topographical Elevation Levels and Eutrophication-Induced Harmful Algal Blooms By Brianna Estocapio

Abstract

Eutrophication and the proliferation of Harmful Algal Blooms (HAB's) have become commonplace in Washington State's freshwater ecosystems in recent years during higher UV exposure months. With the primary contributing variables being nutrient pollution of excess phosphorus (phosphate PO_4) and Nitrogen (ammonia NH_3 / nitrates NO_3^- / nitrites NO_2^-), ultimately such eutrophication-driven hypoxic dead zones[1] can be linked to agricultural sources and incompetent drainage systems. These factors pose the question of how the proportionate relationship between increased nutrient pollution and HAB's could indicate similar proportionality between topographical elevation levels and the growth of phytoplankton and cyanobacteria. Additionally, the water samples taken to test this hypothesis were solely from Washington State freshwater and saltwater sites with relative proximity within or near Seattle. Each site was specific in different elevation levels and water testing strips were compared accordingly.

Introduction

One of the stakeholders in the degradation of United States water quality is the presence of nutrient pollution. This form of aquatic pollution occurs when a surplus of nitrogen (typically nitrate or nitrite) or phosphorus (typically phosphate) enters surface and ground water through upstream freshwater sources such as creeks, natural streams, or rivers. A natural excess supply of such nutrients in bodies of fresh and saltwater may be prevalent because of the erosion of minerals through weathering, human-driven activities like agricultural fertilizer use and improper livestock waste disposal, emission from septic systems and sewage treatment plants, or general storm drain water runoff and incompetent drainage. For comparative measures, the standards that were used as a baseline for these nutrients are outlined by the Environmental Protection Agency (EPA) and Agency for Toxic Substances and Disease Registry (ATSDR) where the aquatic maximum contaminant level (MCL) for nitrites is <1ppm (1mg/L) and <10ppm (10mg/L) for nitrates.[2] Furthermore, nutrient pollution contributes to the development of HAB's through cultural eutrophication. This is the process where humans introduce an abundance of nutrient sources to marine ecosystems so that photosynthetic organisms such as cyanobacteria, noxious phytoplankton, dinoflagellates, and diatoms can easily grow. During the height of a Cyanobacteria Algal Bloom (CyanoHAB), cyanotoxins are heavily produced, and during the height of a noxious phytoplankton's Red Tide the unicellular microalgae also contribute toxins that are fatal to other marine organisms like shellfish and oysters; Additionally, the presence of such algae overgrowth depletes sunlight from other photic zone marine life and raises Potential Hydrogen (pH) levels (≥ 9.2)[3]. In death, HAB's pose a detrimental risk to the aquatic habitat as the decomposition of dead algae exhausts oxygen levels, creating hypoxia. Relating back to nutrient pollution, because phosphorous and nitrogen-rich drainage filters into upstream

waterways that are a convoy for such nutrients, the structuring hypothesis for this paper and accompanying experimentation is that higher geographic elevation levels are correspondent to higher risk of cultural eutrophication and development of HAB's or CyanoHAB's in bodies of water that surround or at the base of those climaxed elevations.

The purpose of this research is to analyze trends in qualitative ecological details and relating pH, nitrite, and nitrate parts per million (ppm) levels while connecting them to topographical elevation data that is approximate to the sample site location. While preceding waterway studies on elevated nutrient levels in Puget Sound and Washington State haven't analyzed a specific connection to elevation, they have outlined trends of change, nitrogen pathways, and the relationship between nutrient concentration and water quality. In a computer-rendered Salish Sea Model from the Washington State Department of Ecology (WSDOE) there are data outputs that suggest the presence of excess phosphorous and nitrogen in the Puget Sound are possibly linked to the contamination of upstream rivers from wastewater treatment plants.[4] Similarly, WSDOE's "Story Map" of nitrogen in Puget Sound records the nitrogen concentrations from anthropogenic activities. From this report wastewater with dissolved inorganic nitrogen in addition to septic system effluent, both contributed >10mg/L of nitrogen into Puget Sound while natural sources only contributed <1mg/L of dissolved inorganic nitrogen.

Methodology

Water samples were taken from 5 freshwater stations within Washington State (Bitter Lake, Green Lake, Ballinger Lake, Pine Lake, and Lake Serene) and 1 saltwater location from Puget Sound (Eagle Harbor). A method of simple probability sampling was exercised as at each of the given locations 3 samples were taken from differing points around the body of water – with the exception being Bitter Lake where 2 samples were collected. Notably, subsequent data was averaged. This ensuring the elimination of research bias, allowing generalizations to be made about the average nutrient levels around the perimeter of such locations. Below, the graphic outlines the connections each of the sample locations has with their elevation levels and GPS coordinates.

Table 1 - Topography Spectrum

High Elevation (440ft-540ft)	Medium Elevation (270ft-400ft)	Low Elevation (0ft-200ft)	Topography Spectrum
------------------------------	--------------------------------	---------------------------	---------------------

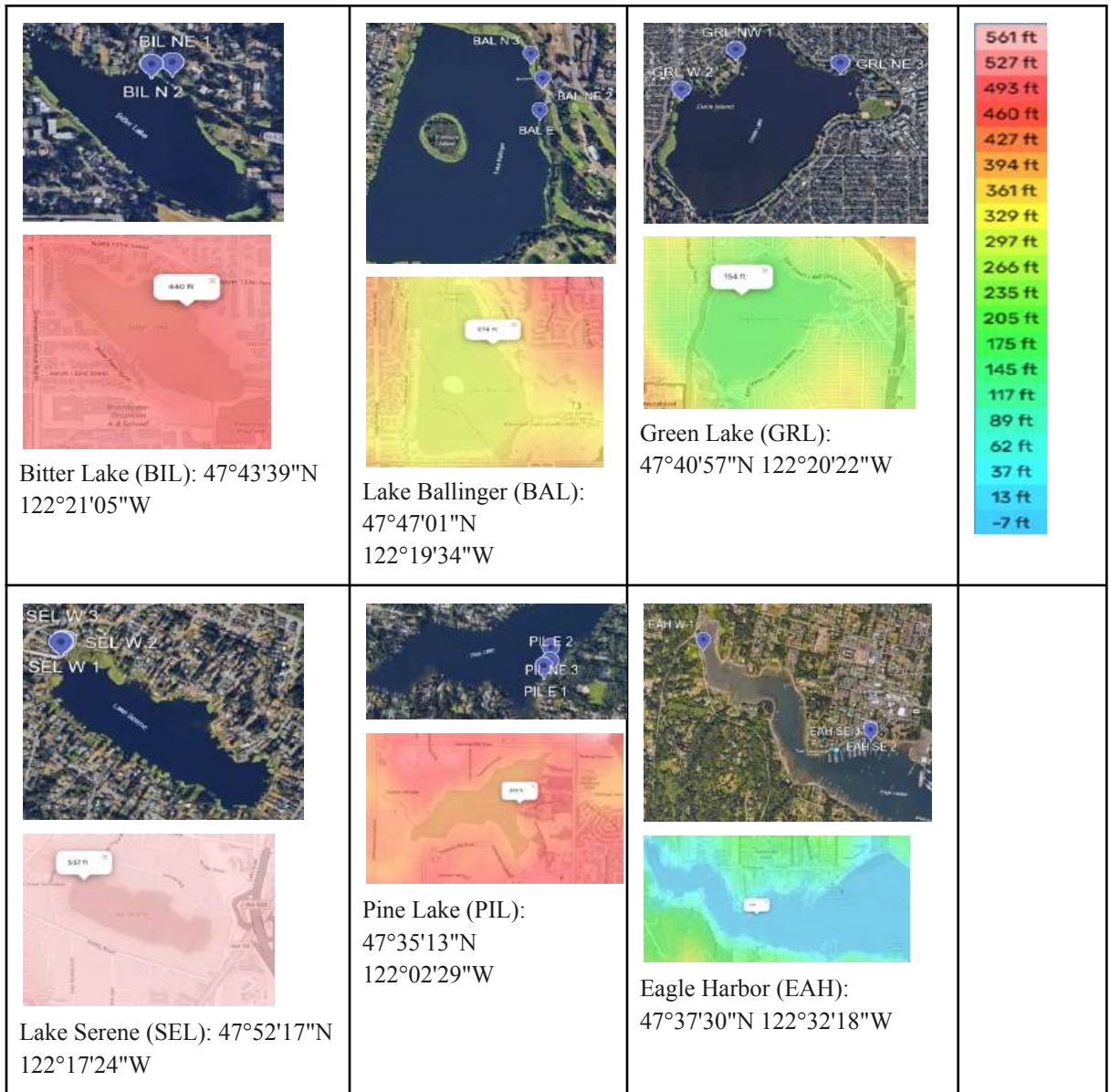


Table 2 – Exact Elevation Values (ft)

High Elevation	Medium Elevation	Low Elevation
Bitter Lake (BIL): 440ft	Lake Ballinger (BAL): 274ft	Green Lake (GRL): 154ft
Serene Lake (SEL): 537ft	Pine Lake (PIL): 393ft	Eagle Harbor (EAH): 0ft

Throughout the experimentation process, the materials used were as follows: 15ml Centrifuge Tubes (sterile) and rack, dial thermometer, glass thermometer, GPS device (digital), Tetra Easy Strips, and API Test Strips.

Procedure


The procedure and protocol taken for equipment was consistent for every sampling and data collection segment. Preceding every location visit, 2 different types of thermometers – both a dial and glass – were calibrated back to the candid starting value of 0°F (~-18°C) by placing them in separate 8oz glasses filled with water and halfway with ice. The thermometers were removed when both read 0°F. As for during a location visit, U.S. Geological Survey (USGS) water quality sampling techniques were applied.[5] For every 3 vertical sections of the body of water chosen, a different centrifuge tube was used by moving the tube vertically from a lower depth moving up toward the surface. After the samples were successfully collected, the temperature in both Celsius and Fahrenheit was collected by almost completely submerging both the dial and glass thermometers in the given locations for >3 minutes. In regards to the sample testing procedure with both the API and Tetra test strips, each strip was dipped directly into the sample, swirled 2 times for 20 seconds, then removed with the pH, NO₃⁻, NO₂⁻, KH (carbonate hardness), and GH (general hardness) pads were faced upward for >30 seconds before reading the NO₃⁻, NO₂⁻, or pH values.



Criteria was also attributed to analyzing the qualitative data from ecological signs upon arrival at each location. For this collection of qualitative data, characteristic standards were taken from the Washington State Department of Health (DOH)[6] and pertained to determining how much sunlight the sample site received and therefore how much surrounding foliage there was. If algal mats or foam were present on the surface, green or opaque discoloration was present as a result of cyanopigments (Blue-Green Algae), and the presence of dead marine life. After the collection process, samples were stored in a 34°F container and covered completely by a dark cloth.





Results

The data for this experimentation is two-fold, with the qualitative being aquatic characteristics of the sample locations and the quantitative being the Tetra and API test strip values for pH, NO₃⁻, NO₂⁻, KH, GH, and Cl (Chlorine only applicable to Tetra).

Table 3 – Qualitative and Temperature Values[7]

#	Location	Temperature (°F)	Water Type	UV Exposure	Ecological Characteristics	Visuals
1	BIL NE	≈70°F	Freshwater	Part Shade	Consolidation of some algae mats, on-water lily pad foliage, limited green discoloration	
2	BIL N	≈70°F	Freshwater	Full Sun	Mostly clear, no foliage	

1	GRL NW	≈72°F	Freshwater	Full Shade	Medium to dense surrounding foliage, green organic matter clots, hydrilla noxious aquatic weed growth	
2	GRL W		Freshwater	Part Shade	Dense surrounding foliage, no discoloration, clear shoreline	

3	GRL NE	≈80°F	Freshwater	Full Sun	No surrounding foliage, brown discoloration	
1	BAL E	≈65°F	Freshwater	Full Sun	Sparse foliage, very clear shoreline	
2	BAL NE	≈65°F	Freshwater	Part Shade	Dense surrounding Blue Lyme Grass, lily pad growth, iridescent gasoline film on shore	
3	BAL N	≈60°F	Freshwater	Full Sun	Sparse low foliage, slight algal matting, large lily pad growth	
1	PIL E	≈72°F	Freshwater	Full Sun	Semi-dense surrounding foliage, clear shoreline water	
2	PIL E	≈71°F	Freshwater	Full Sun	No surrounding foliage, noxious aquatic weed growth (water lilies)	




3	PIL NE	≈73°F	Freshwater	Part Shade	Dense surrounding foliage, dense noxious aquatic weed growth, no discoloration	
1 2 3	SEL W	≈72°F	Freshwater	Full Sun	Minimal surrounding foliage, dense lily pad growth, visible marine life	
1	EAH W	≈77°F	Saltwater	Full Sun	Dense surrounding foliage, extremely opaque water, shoreline clay	
2 3	EAH W	≈62°F	Saltwater	Full Sun	No surrounding foliage, mostly clear shoreline	

Table 4 – Quantitative API Test Strips and Tetra Easy Strips (Cl Values)

Sample Number	Location	pH	NO ₂ (Nitrite) ppm (mg/L)	NO ₃ (Nitrate) ppm (mg/L)	KH (Carbonate Hardness) ppm (mg/L)	GH (General Hardness) ppm (mg/L)	Cl (Chlorine) ppm (mg/L)
1	BIL NE	7.5	0.5	20	40	30	0.5
2	BIL N	6.5	0	0	0	30	0
1	GRL NW	8.0	1	20	40	120	0
2	GRL W	8.0	0.5	0	40	60	0
3	GRL NE	8.0	0	0	40	60	0
1	BAL E	7.5	0	0	40	60	0
2	BAL NE	7.5	0	0	40	60	0
3	BAL N	7.5	0	20	80	120	0
1	PIL E	6.5	0	0	0	30	0
2	PIL E	6.5	0	0	0	0	0
3	PIL NE	7.0	0	0	0	0	0.5
1	SELW	7.0	0	0	40	30	N/A[8]
2	SEL W	7.0	0	0	40	0	N/A
3	SEL W	0.5	0	0	40	0	0
1	EAH W	7.5	0	0	120	180	N/A
2	EAH SE	7.5	0	0	80	180	N/A
3	EAH SE	7.5	0	0	80	180	N/A

Table 5 – Averaged Quantitative data

Location	pH	NO ₂ ⁻ (Nitrite) ppm (mg/L)	NO ₃ ⁻ (Nitrate) ppm (mg/L)	KH (Carbonate Hardness) ppm (mg/L)	GH (General Hardness) ppm (mg/L)	Cl (Chlorine) ppm (mg/L)
BIL	7.0	0.2	10	20	30	0.2
GRL	8.0	0.5	10	40	80	0
BAL	7.5	0	10	53.3	80	0
PIL	6.6	0	0	0	10	0.2
SEL	4.8	0	0	40	10	N/A
EAH	7.5	0	0	93.3	180	N/A

Analysis and Discussion

Note: As KH, GH, and Cl are not primary indicators of phytoplankton growth or eutrophication. pH and inorganic NO₃⁻ or NO₂⁻ will be considered when analyzing the relationship between elevation and nutrient pollution.

Despite misaligning with the rudimentary conclusion that indicates higher elevation creates a higher risk for eutrophication, the pH, NO₃⁻, NO₂⁻, and ecological observations were found to relate to topographical elevation through the criteria of surrounding urbanization and difference in elevation between encircled terrain and the given body of water. When applying these criteria, in order to gauge the extent of surrounding urbanization near a sample site, the United States Census Bureau demographic data map[9] was utilized to situate urbanization as a first consideration. The highest population levels were found within or just north of Seattle. Green Lake (GRL) had an approximate 5,355 surrounding population and Lake Ballinger (BAL) had an approximate 5,693 surrounding population. These statistics align wholly with the quantitative pH results which show GRL and BAL to also have the highest average pH levels of 8.0 and 7.0. Because both nutrients such as nitrogen and a phytoplankton abundance such as HAB's raise pH, the higher average for GRL and BAL indicates more nutrients that could be sourced from the greater surrounding population with urbanization contributing to an elevated runoff percentage. In addition to this, the stark pH fluctuations that occur from such an excess of nutrients and subsequent photosynthesis (from benefiting algae) were accounted for as samples were taken in the afternoon when pH levels were at their highest. In order to connect the GRL and BAL pH levels and generalized urbanization relationship back to elevation levels, the difference in feet between the elevation of the lakes and surrounding elevation must be considered due to nutrient-permeated runoff and drainage is likely to flow toward freshwater lakes located on steeper terrain. GRL had the highest difference of elevation at 199 ft. While the lake itself had an elevation of 154 ft, the surrounding area had an average elevation of 343 ft. The second highest difference, of 81 ft, was observed at BAL. Again, the elevation data aligns with the qualitative data for sample site locations with the highest pH, therefore a risk of developing HAB's or CyanoHAB's is possible

Because of such imperative conclusions, the connection can be made that both variables of urbanization percentage and elevation difference between sample sites and surroundings have a proportional relationship with subsequent nutrient pollution, eutrophication, and potentially harmful phytoplankton (cyanobacteria, noxious phytoplankton, dinoflagellate, or diatom) blooms.

This verdict also matches the quantitative data for low pH, population levels, and elevation difference for Pine Lake and Lake Serene (PIL and SEL), and medium pH, population levels, and elevation difference for Eagle Harbor and Bitter Lake (EAH and BIL). However, the qualitative data is slightly more difficult to categorize. While there were visible sections of noxious aquatic weed growth and algal matting at Green Lake in both high and low UV exposure areas (and corresponding temperature), Lake Ballinger ultimately showed less conspicuous aquatic attributes and instead had prominent drainage attributes of drainage domes directly on the shoreline which filtered into the lake itself. This offsetting of different attributes were prevalent for other sample locations as well, making the signs of nutrient pollution less straightforward.

In regards to quantitative data for inorganic nitrate and nitrite levels, the criteria of population size and elevation difference is applicable. The primary difference between average nitrogen statistics and pH statistics were the absence of a “medium level” categorization for nitrogen parts per million (ppm). Despite this, higher NO_3^- ppm was found at GRL, BIL, and BAL while higher NO_2^- ppm was found at GRL and BIL. Lower NO_3^- ppm were observed at PIL, SEL, and EAH while lower NO_2^- ppm were observed at PIL, SEL, EAH, and BAL (See table 5). Once again the varying levels of NO_3^- and NO_2^- match the percentage of population size and elevation difference as outlined by the pH observations. . Consequently, NO_3^- and NO_2^- can be later converted into ammonium (NH_4^+) which is the most bioavailable for cyanobacteria or other HAB phytoplankton to use as biomass. Higher ppm levels at GRL, BIL, and BAL indicate an inflated probability that nutrient pollution is prevalent within the tested freshwater lake systems, and therefore a possibility for HAB's to develop is emanate.

Now, focusing on the lower inorganic nitrate and nitrite levels attributed to PIL, SEL, and EAH, the surrounding population size for each of the respective locations was below 3,810 (SEL/EAH being in the “low” category and PIL being in the “medium” category), each had a much lower elevation difference of >425 ft. These statistics align with the subordinate NO_2^- and NO_3^- ppm while additionally matching the qualitative data. While noxious aquatic weed growth (such as water lilies) was prevalent at SEL, EAH, and PIL, ultimately there was little to no discoloration or algal matting in the water and the shoreline for each location was relatively clear.

Conclusion

Considering that when surrounding population influence and elevation difference is proportionally lower, the nitrogen and pH levels are also lower, a correlation of a directly proportional relationship between elevation difference, urbanization, and eutrophication of

surface water (greater possibility for HAB development) can be derived. Photosynthesizing organisms and toxin-producing phytoplankton thrive in higher nitrogen concentrations and create more drastic fluctuations in pH. These attributes were seen prominently at the sample locations with the highest amount of surrounding urbanization and population in addition to having the greatest elevation difference between surrounding terrain and the given body of water.

Limitations

The primary limitations of the preceding experimentation included the lack of access to phosphorus sensors and electrochemical probes for dissolved oxygen. The data from such equipment could have further corroborated my hypothesis, offered support, or modified conclusions based on the pH levels and measurements of inorganic nitrate or nitrite levels. Phosphorous, a large component of nutrient pollution and elevated phytoplankton or dinoflagellate growth, heavily depletes dissolved oxygen levels creating hypoxic “dead zones”. Measurements of dissolved oxygen and phosphorus could verify which locations were susceptible to eutrophication. Additionally, the use of a light microscope may have aided the identification of which sample sites had the most prevalent phytoplankton or dinoflagellate growth. which sites could develop HAB’s, CyanoHAB’s, or non-toxic algal blooms with increased UV exposure and nutrients.

Works Cited

- Agency for Toxic Substances and Disease Registry . (2021, February 9). Nitrate/Nitrite Toxicity: What Are U.S. Standards and Regulations for Nitrates and Nitrites Exposure? | Environmental Medicine | ATSDR. [Www.atsdr.cdc.gov](http://www.atsdr.cdc.gov); ATSDR.
<https://www.atsdr.cdc.gov/csem/nitrate-nitrite/standards.html#:~:text=EPA%20has%20set%20an%20enforceable%20standard%20called%20a>
- Algae Research Supply, HUBER, M., & BLAHA-ROBINSON, K. (2017). Algae Culture and pH. Algae Research Supply. <https://algaeresearchsupply.com/pages/algae-culture-and-ph>
- Causes and Ecosystem Impacts | Harmful Algal Blooms | CDC. (2022, October 4). [Www.cdc.gov](http://www.cdc.gov).
<https://www.cdc.gov/habs/environment.html#:~:text=However%2C%20three%20main%20types%20of%20phytoplankton%20cause%20most>
- Factsheet on Water Quality Parameters. U.S. Environmental Protection , July 2021, www.epa.gov/system/files/documents/2021-07/parameter-factsheet_nutrients.pdf.
- Figueroa-Kaminsky, C. (2014). Salish Sea modeling - Washington State Department of Ecology. [Wa.gov](http://wa.gov).
<https://ecology.wa.gov/Research-Data/Data-resources/Models-spreadsheets/Modeling-the-environment/Salish-Sea-modeling>
- Fouts, M., Splitz, M., Anaya, N., & Karnae, S. (n.d.). Effect of Nitrates on Algae Bloom Formation – 10th Sustainability Symposium. Department of Environmental Engineering and Earth Science; Wilkes University . Retrieved December 26, 2023, from <https://sustainabilitysymposium.scholar.bucknell.edu/2021/04/22/effect-of-nitrates-on-algae-bloom-formation/>
- Gann, Eric R., et al. “Elevated pH Conditions Associated with Microcystis Spp.. Blooms Decrease Viability of the Cultured Diatom *Fragilaria Crotonensis* and Natural Diatoms in Lake Erie.” Edited by Brittany N. Zepernick, *Frontiers*, 20 Jan. 2021, www.frontiersin.org/articles/10.3389/fmicb.2021.598736/full.
- Gilliom, R., & Bortleson, G. (1983, January 1). Relationships between water quality and phosphorus concentrations for lakes of the Puget Sound region, Washington | U.S. Geological Survey. [Www.usgs.gov](http://www.usgs.gov); United States Geological Survey.
<https://www.usgs.gov/publications/relationships-between-water-quality-and-phosphorus-concentrations-lakes-puget-sound#:~:text=The%2078%20lakes%20evaluated%20in%20the%20study%20had>
- Great Lakes HAB's Collaboratory. (n.d.). How Does Nitrogen Affect Harmful Algal Blooms? (p. 3). Retrieved December 27, 2023, from <https://www.glc.org/wp-content/uploads/HABS-Role-of-Nitrogen-20170912.pdf#:~:text=Nitrate%20and%20nitrite%20%28NO3-%2FNO2%20-%29%20must%20be%20actively>
- Harmful Algal Blooms. (2019). [Noaa.gov](http://noaa.gov); National Oceanic and Atmospheric Administration .
<https://oceanservice.noaa.gov/hazards/hab/>
- Johnson, A. (2022, August 29). How does algae affect the pH of water? *ScienceOxygen*.
<https://scienceoxygen.com/how-does-algae-affect-the-ph-of-water/>

- Lindquist, S., & Goldoft, M. (2021). Harmful Algal Blooms. In *A Monthly Bulletin on Epidemiology and Public Health Practice in Washington* (p. 5). Washington State Department of Health .
<https://doh.wa.gov/sites/default/files/legacy/Documents/5100/420-002-epitrends2021-06.pdf>
- McCarthy, S., Mohamedali, T., & Cracknell, P. (n.d.). Story Map Series. [Waecy.maps.arcgis.com](https://waecy.maps.arcgis.com); State of Washington Department of Ecology . Retrieved December 26, 2023, from <https://waecy.maps.arcgis.com/apps/MapSeries/index.html?appid=907dd54271f44aa0b1f08efd7efc4e30>
- Ruckelshaus, M., & McClure, M. (2007, January). Harmful algal blooms in Puget Sound | Encyclopedia of Puget Sound. www.eopugetsound.org.
<https://www.eopugetsound.org/articles/harmful-algal-blooms-puget-sound>
- Smith, B. (n.d.). Nitrate & Nitrite Rule for Transient Non-Community Water Systems. In Environmental Protection Agency (p. 22). Retrieved December 26, 2023, from https://www.epa.gov/sites/default/files/2018-05/documents/nitrate_and_nitrite_rule_training_presentation.pdf
- Topographic Maps. (n.d.). Seattle Topographic Map, Elevation, Terrain. King County Maps. Retrieved December 27, 2023, from <https://en-us.topographic-map.com/map-plz4/Seattle/?center=47.63798%2C-122.48056&lock=12%2C-2%2C171&zoom=14&popup=47.62008%2C-122.52243>
- United States Geological Survey . (2018, June 8). Water Quality Sampling Techniques | U.S. Geological Survey. www.usgs.gov; Water Science School.
<https://www.usgs.gov/special-topics/water-science-school/science/water-quality-sampling-techniques>
- US Census Bureau. (2019, March 20). 2020 Census Demographic Data Viewer. [Census.gov](https://www.census.gov).
<https://www.census.gov/programs-surveys/geography/data/interactive-maps.html>
- Washington State Toxic Algae. (2012). www.nwtoxicalgae.org; Washington State Department of Health. <https://www.nwtoxicalgae.org/HistoricalCharts.aspx>
- Yarimizu, K., Fujiyoshi, S., Kawai, M., Norambuena-Subiabre, L., Cascales, E.-K., Rilling, J.-I., Vilugrón, J., Cameron, H., Vergara, K., Morón-López, J., Acuña, J. J., Gajardo, G., Espinoza-González, O., Guzmán, L., Jorquera, M. A., Nagai, S., Pizarro, G., Riquelme, C., Ueki, S., & Maruyama, F. (2020). Protocols for Monitoring Harmful Algal Blooms for Sustainable Aquaculture and Coastal Fisheries in Chile. *International Journal of Environmental Research and Public Health*, 17(20), 7642.
<https://doi.org/10.3390/ijerph17207642>

Duolingo Hindi and Short Term Memory Retention By Sai Smriti Saiganesh

Abstract

The research focus is evaluating Duolingo's ability to instill Hindi Characters in the short term memories of non-native Hindi speakers. The problem addressed in this paper is that non-native Hindi speakers are blindly gravitating towards technological resources without deeply understanding the benefits or usefulness of that particular resource; especially Duolingo since it is among the most popular language learning apps. This paper continues the investigation of Sandeep Saini and Vineet Sahula from Language Learnability Analysis of Hindi: A Comparison with Ideal and Constrained Learning Approaches. This paper was inspired by the problem discussed by Sandeep Saini and Vineet Sahula in the work mentioned above; they discuss how new language learners gravitate towards language learning resources without knowing their constraints. Furthermore, Saini and Sahula stress the importance of finding an efficient method to learn languages, especially the Hindi language. The method used in my study was the evaluative research method. Results showed Duolingo was not efficiently able to instill the Hindi characters in the short-term memories of non-native speakers overall. This disagrees with most current understandings, because a majority of the current research shows that Duolingo is a reliable and beneficial app to those looking to acquire a new language. Ultimately, it was found that Duolingo is unpredictable and unreliable for non-native speakers looking to improve their short-term memorization of the Hindi language. Future investigations should explore other language learning apps' abilities to teach Hindi and Duolingo's ability to instill Hindi characters in the long-term memories of non-native speakers, while expanding the sample size.

Literature Review

Almost everywhere we go, we view a world surrounded by technology in a variety of aspects. Technology has now become a substitute for some people's jobs and commonalities. Teaching, one of the most well-known and resourceful jobs, has been impacted by technology a great deal as well. The integration of technological components into learning has been termed: E-learning. E-learning is educating through electronic media and essentially the utilization of the internet and its resources to gain knowledge. E-learning combats in-person learning in the current day while learners and educators weigh the benefits and losses of education through technological means (Stecula and Wolniak). One of the greatest aspects of learning that is greatly assisted by technology is learning a language, which is termed: linguistics, or the science regarding the learning of a language. With its unique characters and complexities, Hindi has become a language that learners are gravitating towards to receive aid through the convenience of technology. The origins of this language are rooted in India and its diverse culture. This paper will focus on the language learning app, Duolingo, and the extent to which it is able to ensure the short term memory retention of the Hindi language characters in non-native Hindi speakers at a large urban high school in northwestern Washington state.

From a historical context, e-learning has been a relatively new addition to the lives of many, at least in the last few decades or so. Paul Nicholson of *A History of E-Learning* describes e-learning as “opportunities being multiplied as they are seized” (Nicholson 2). He continues the discussion about the history of e-learning by introducing the idea that educators are essentially relying on technology and computers to help enhance the quality of learning for their students as opposed to the past in which they did not have these resources available to them (Nicholson 2). E-learning serves as a potentially beneficial resource to those who intend to further their knowledge to further levels than ever before. Adding on, E-learning, which is also termed as distance learning, started its rapid growth in the late 1900s alongside the advancements of the online technological revolution. E-learning took on many forms in the 1900s such as parcel post, radio, television, and eventually online education (Kentnor 4). In this way, E-learning has kept up with the public even through long distance barriers. Knowing that E-learning has been growing alongside advancements in technology, it is safe to say that E-learning will only continue to grow regularly and play a more fundamental role in our lives as technology continues to evolve as well. This direct relationship helps to demonstrate how the historical importance of E-learning still holds its ties in today’s more technologically advanced society. It has made more opportunities available for individuals to receive education and expanded the general public’s knowledge as well.

However, this being said, other learners have found e-learning applications, like Duolingo, to be more detrimental to their learning than beneficial. One aspect is that the gamification, or utilization of game elements in non game contexts to engage users (The Definite Definition) of Duolingo poses a threat to an individual's education. It is said by some that Duolingo leads to gamification misuse in which users focus more on the game rather than the actual education associated with it. This wastes users’ time and negatively impacts their educational performance (Mogavi, et al.). In a study examining in person versus online learning, in terms of utility versus social experience, the majority of the students expressed that in person education was superior to e-learning because it was easier to communicate with the professor (66%), concentrate for longer periods of time (73%), and understand the material (81%) (Photopoulos). These trends follow learners to Duolingo’s Hindi curriculum as well. Some individuals learning the Hindi language feel as though Duolingo has strayed far from the authenticity of the Hindi language (Gordon). Adding on, some feel that a prominent issue with Duolingo’s Hindi curriculum was the difference between the phonetic letters for the letters in their tools and the original Hindi letters. That along with the lack of explanations behind the grammar of the language left learners confused and negatively impacted their education (Gordon). With reviews like this, it becomes clear that Duolingo has negative impacts on the learning of certain groups of individuals. It is important to recognize these cons as they present room for improvement in e-learning today. Therefore, Duolingo can be thought insufficient in its ability to help its learners in some instances and the e-learning community may need more support from additional resources.

As non-native Hindi speakers gravitate towards online language learning apps, such as Duolingo, to learn the Hindi language, they depend heavily on Duolingo's ability to ensure their memory retention of the characters, tones, and vocabulary. This is seen through Duolingo being the world's number one language learning app on the app store and google play app store (Duolingo's Gamified Success) (See Appendix A) . Language learning involves a variety of aspects such as memorization, comprehension, etc. Duolingo can be seen to be beneficial in aiding with that, through their research study in which they conducted a cross-sectional analysis about two different language courses that they offer. Duolingo's Spanish and French courses improved the reading, comprehension, and memory retention of its users at the end of the course substantially (Jiang 2022). Duolingo also assessed their ability to insure the speaking skills of their users through a data collection and data analysis of their Spanish and French course users' data. The study showed that Duolingo significantly improved the speaking skills of its Spanish and French users a great deal through the parallel use of the speaking and listening components available to the users in the app (Jiang 2021). Overall, it can be seen that Duolingo is well-rounded in a variety of aspects in terms of the linguistics of Spanish and French in its users, but they do not address the linguistics behind their Hindi course or their ability to aid the memory retention of the Hindi language in any of their published research papers. The lack of a study on the short-term memory retention abilities of Duolingo Hindi is also under-researched in the education and technological fields as well.

Hindi is also a language with deep cultural significance. A little under half the population of India claim Hindi as their mother tongue and even more can understand it proficiently. The Hindi language contributed significantly to other Indian languages and can potentially be referred to as the mother tongue of all the Indian languages because of Hindi's religious significance and its original literatures (Sathyanarayanan 1). From a cultural perspective, Hindi's cultural origins and contributions are rooted all throughout India. Its great popularity demonstrates the importance of its research as such a large population of people associate with it. Another view along the lines of culture, is that "immersion in the language environment online can actually lead to a better understanding of the culture itself" (Edmundson xi). The view aligns technology and culture in parallel when discussing linguistics. Through utilizing e-learning and technological resources, language learners are enabled to feel a deeper connection to the language as they experience and learn the culture behind it as well. It's important to address culture when teaching a language, because it enables language learners to be able to communicate with a higher proficiency (Kramsch 218). Through having a general idea of culture, learners can understand how worldly views differ and how languages can accommodate those ideological differences.

My study will combine data collection and data analysis from non-native Hindi speakers at a large urban high school in northwestern Washington state to assess the ability of Duolingo to ensure memory retention of a variety of Hindi characters. The significance of this study is to address the gap that exists in-terms of the lack of a focused study on the Duolingo Hindi curriculum in terms of its short term memory retention abilities and non-native learners. My

study will help these non-native Hindi learners navigate the world of linguistics from an e-learning standpoint and ensure they gravitate towards resources that are beneficial to their learning of the Hindi language instead of turning to e-learning applications, like Duolingo, with a blind eye and no background knowledge as to how and if it aids their memory retention and proficiency of the Hindi language.

Methodology

The method type used in this study was evaluative research. This method essentially aims to assess a specific problem and increase the usability and productivity associated with the product/ application with that problem. It also evaluates its reliability of that product to the public (Suchman 83) and suggests solutions and implications based on the results produced by the study. Through using this method, I will be evaluating Duolingo's effectiveness in instilling Hindi characters in the short term memories of non native Hindi speakers. By utilizing the evaluative research method, this study will be able to produce results that will enable participants to know if Duolingo is a good resource for them to learn the Hindi language.

The project participants will be students at a large urban high school in northwestern Washington state that are also non-native Hindi speakers. All participants in this study were 11th graders that attended this respective high school. Due to this high school having a relatively large population of students, the decision to focus on 11th graders was made. There are roughly 430 juniors at this high school and approximately 40 percent of this institution is Asian. Taking into account that only non-native Hindi speakers were being evaluated for this study and that Hindi is a language of Asian (specifically Indian) descent, I subtracted out this 40 percent from 100 percent of the total population and got 60 percent which represented the approximate percentage of non-asian students at this school. Even though not all Asians speak Hindi, this math was done to eliminate further chances of skewing my data with Hindi speakers, since Hindi does fall in the category of Asian languages. Then, the remaining 60 percent of the 11th grade population was calculated (430×0.60) to be 258. Finally, due to time and resource constraints, 10% of the population size was taken (258×0.10) and produced the sample size of 25.8 which was rounded down to 25 participants. This is how the sample size was calculated for this research study.

The participants will be completely informed of everything their participation entails before the start of their involvement in this study and this will be followed by a required informed consent form for the participants to officially commit to their participation. Each of the 25 participants were required to sign physical copies of this informed consent form prior to their participation in this study (see Appendix B). This form states that they are allowed to decline participation at any time for any reason. It also states that participants' personal data/information, such as names, will not be published to the public view. This is to protect the privacy of each participant.

Each individual who participated in this study followed the same routine. Over a course of two weeks, May 5th to May 17th (not including Sunday), each participant was tasked to learn three Hindi characters a day. The two week timeline is to represent two trials and to increase the

validity of my results. This timeline is also to test the short-term memory component of Duolingo's Hindi curriculum as it only requires participants to retain information for a week or less before they are assessed on it. The 15 characters learned in the first week (in order) were: क, ख, ग, घ, च, छ, ज, झ, ट, ठ, ड, ढ, त, थ, and ड. Additionally, the 15 characters learned in the second week (in order) were: ध, न, प, फ, ब, भ, म, य, र, ल, व, श, ष, स, and ह. This introduced a total of 30 Hindi characters to the participants' short term memories. This learning was done through daily zoom calls with me, the researcher, and utilizing Duolingo's own curriculum. This means that the questions appeared exactly as they would on the Duolingo app (see Appendix C). For each Hindi character, the participant was allowed to hear the sound of the character play and then had to choose the corresponding English phonetic letters that they felt best matched it. They were immediately notified on the screen about their accuracy and were told what the correct answer was if they chose the wrong answer. For each week, Monday through Friday consisted of lessons (learning three different characters per day) and Saturday was a cumulative test of every character the participants had learned in that respective week. This means that the cumulative tests were 15 questions long, representing the 15 characters they learned that week. The test questions looked and worked exactly like the practice questions; however, in the test, participants were not given the option to listen to the pronunciation of the characters. They had to associate the image of the character to the English phonetic letters without the hearing sound. The reason for this being that Duolingo's tests work the same way in which users cannot hear the sound of the character when deciding its phonetic spelling. This situation challenged the participants to recall the association between the Hindi characters and their phonetic spellings which were taught to them in the practice sessions and embedded in their short term memories.

This research study does not involve any major ethical concerns to the environment, participants, the institution of the participants, or I (the researcher). The only possible ethical concern could be participants getting overwhelmed learning new material. However, they are informed that they can decline participation at any time in the consent form. The original raw data (before renaming participants as "participant #") collected will include names, but not other personal information. When the study is officially published, no names will be disclosed, therefore there will not be any privacy concerns. The recording of participants names in the original raw data is done to ensure that every participant learns every character. This is also for the researcher to stay organized with their participants and their individual data. The only data that will be collected in my study, will be the accuracies of each of my participants through the material each week over the two week period. All the data will be recorded in a google spreadsheet and will then be destroyed after the study comes to an end.

Findings

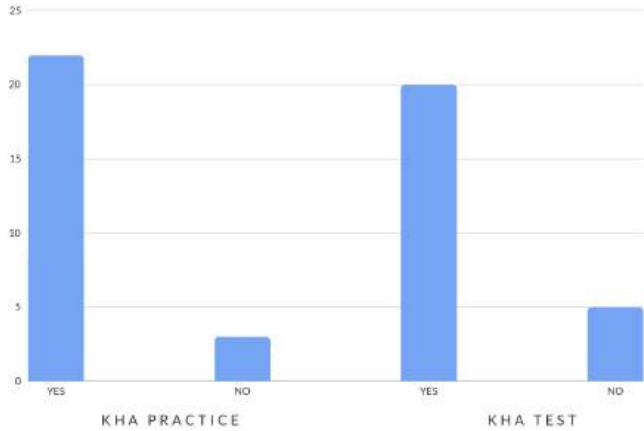
My two week research process primarily consisted of zoom calling participants, having them learn three new Hindi characters per day, and taking an accumulative test at the end of the week (two separate tests of the 15 characters they learned that week). Throughout this, all the participants' accuracy data was put into Google Sheets (separated for week one and two). Each

participant was labeled “participant #” to protect the personal information of the participants. This data was also my raw data that I eventually analyzed on a deeper level to lead to my conclusion. The raw data collected through this procedure is represented in the tables directly below. Each column represents the accuracies for one character. For example, column 1 of February 5th, represents the accuracies of the क, or kha, character. The order in which the columns are organized is the same as the chronological order that the characters were taught in. This order was mentioned above (क, ख, ग, घ, च, छ, ज, झ, ट, ठ, ड, ढ, त, थ, ड for week one and ध, न, प, फ, ब, भ, म, य, र, ल, व, श, ष, स, ह for week two). In these tables “Y” represents “yes”, a correct answer, and “N” represents “no” for an incorrect answer:

week 1	2/5	2/6	2/7	2/8	2/9	2/9
participant 1	YNN	YYN	NYN	YYN	YYN	YYNYNYNYNYNYNY
participant 2	YNY	NNY	NNY	YNY	YNN	YNYNYNNYNNYNNYY
participant 3	YYY	YNY	YNY	NNN	NNY	YYYNYNYNNYNNYY
participant 4	YNY	YYY	NNY	YNN	YYN	NNYNNYNNYNNYNN
participant 5	NYN	NYN	YNY	NYN	NNY	YNNNYNYNYNNYNY
participant 6	NNY	YYN	NNY	YNY	YNY	YNYNYNYNNYNYNY
participant 7	YYY	NNY	NNN	YYN	NYN	NYNYNYNNYNYNYNY
participant 8	YYY	YNY	YNN	YYY	YYN	NNYNYNYNYNNNNY
participant 9	YNN	YYN	YYN	YNY	YYY	YYNNYNYNYNNNN
participant 10	NYN	NNY	YYY	YNY	NNY	YYYNYNNYNNYNN
participant 11	YNN	NNN	NYN	YYN	NNN	NYNNYNYNYNNYNY
participant 12	YYN	YNN	YNY	YYN	YYY	NNNYNYNYNYNNY
participant 13	YYN	NYN	NYN	NNY	NYN	YNNNYNYNYNNNNYY
participant 14	YYN	YYN	YNY	YNY	NYY	YYYNYNYNYNYNNN
participant 15	NYY	YNY	YNY	NYY	YYN	YNNYNYNYNYNYNY
participant 16	NNN	NNN	NNN	YNN	YNY	YNYNNYNYNYNYNY
participant 17	YYN	YNN	YNY	YNY	YNY	YNYNYNNNNYNYNY
participant 18	YNY	YNY	YNY	YYN	YNY	YNYNYNYNNNNYNY
participant 19	YYN	YNN	YNN	NNY	YYN	NNYNYNYNYNYNYYY
participant 20	YYY	YNY	YNY	YNY	YNY	YNNYNYNYNYNYNY
participant 21	YNN	NYY	YNY	YNY	YNN	YNNYNYNNNNYNYNN
participant 22	YNN	YYN	YYY	NYN	YYN	NYNYNYNYNYNNNN
participant 23	YYN	NNN	YNY	NNY	YNY	YYYNNNYNYNNNNY
participant 24	YNY	YNN	NNY	YNY	YNY	NYNNNNYNYNNNNYY
participant 25	YYY	YYN	YNY	NNN	YNY	YNYNYNYNYNYNN

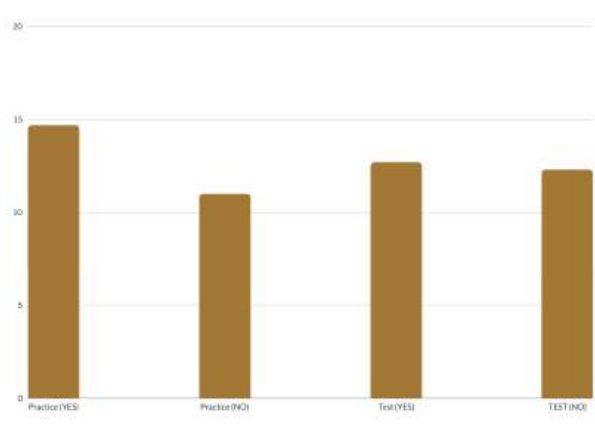
week 2	2/12	2/13	2/14	2/15	2/16	2/17
participant 1	Y N N	Y N Y	Y N Y	Y N N	N Y Y	N N Y N N Y N Y N Y N Y Y
participant 2	Y N Y	Y Y Y	Y N N	N Y Y	Y Y N	N N N Y N N N N N Y N Y N Y Y
participant 3	N Y Y	Y N N	N N Y	N N N	Y Y Y	N N N Y Y N Y N Y N Y N N N
participant 4	N Y Y	Y N N	Y Y Y	Y N Y	N N Y	N Y N N Y Y N Y N Y N N N N N
participant 5	Y N Y	Y Y N	N Y Y	Y N Y	Y Y N	N N Y Y N N Y N Y N N Y N Y N Y
participant 6	Y Y Y	Y Y N	Y N Y	Y N N	Y Y Y	N N Y N Y N Y N N N N Y N Y N
participant 7	N Y Y	N Y Y	Y N Y	Y Y Y	N Y Y	N N N Y N N Y N Y Y Y N N N N
participant 8	Y Y Y	Y N Y	Y N Y	N Y N	N Y Y	Y N N N N Y N N Y Y Y Y N Y N
participant 9	Y N Y	Y N Y	Y Y N	Y Y Y	N Y Y	Y Y Y Y N Y N Y N Y N N N N N
participant 10	Y Y N	Y Y Y	N N Y	N N N	Y N Y	Y N Y N Y N Y N N N N N Y Y Y
participant 11	Y N Y	N N N	Y N Y	N Y N	Y N Y	Y N Y Y N Y N Y N Y N N N Y
participant 12	N Y N	N Y Y	Y N N	N Y Y	Y N Y	Y N N Y N Y N Y Y Y N Y N N
participant 13	Y Y Y	N N Y	N Y Y	Y Y Y	Y Y N	Y Y N N N Y N N N N Y N Y Y Y
participant 14	Y N Y	Y N N	N Y N	Y N N	N N Y	Y Y N Y N Y N N N N N Y N N
participant 15	Y N N	Y Y Y	Y N Y	N Y N	Y N Y	Y N Y Y Y Y Y N Y N N N N N N
participant 16	N Y Y	Y N N	Y Y Y	N N N	Y N N	N Y Y Y N N Y Y Y Y N N Y N Y
participant 17	Y N N	Y Y Y	Y Y N	Y Y Y	Y N Y	Y Y N N Y N Y N N N Y N N N
participant 18	Y N Y	Y N N	Y N N	N Y Y	N N N	N Y N N Y Y N N N Y N Y Y Y
participant 19	Y N N	Y Y Y	Y N N	Y N N	Y N Y	Y Y Y N Y Y Y Y Y N Y N Y N Y
participant 20	N N Y	Y N N	Y Y N	Y Y Y	N Y Y	Y N N N Y N Y N N N N Y N N
participant 21	N Y Y	Y N Y	N N Y	N N Y	N Y N	N Y Y N Y N N N Y N Y N N N Y
participant 22	Y N N	Y N Y	Y N Y	N Y Y	Y Y Y	N Y N N Y N Y Y Y N Y N Y N N
participant 23	Y Y Y	Y Y Y	Y Y N	N Y N	N N Y	N Y N Y Y Y N N N N Y N Y N Y
participant 24	Y Y N	Y Y Y	N Y N	N N N	Y Y Y	Y Y Y Y N N N Y Y Y Y Y Y N
participant 25	N Y Y	N Y N	Y N Y	Y Y N	Y N Y	N Y Y N Y Y Y Y N Y N Y Y Y Y

The next step was to analyze and interpret this data. To begin, the accuracy trend for one singular character was analyzed. This was the character **ख**, or kha. This character was chosen for a specific trend analysis through random selection. To conduct this random selection, all 30 characters were put into a random result generator and the program’s algorithm was tasked in deciding the character. By looking at a random sample of the overall data, trends between the accuracies of one singular character and overall data can be seen. The kha accuracy trends were as follows:

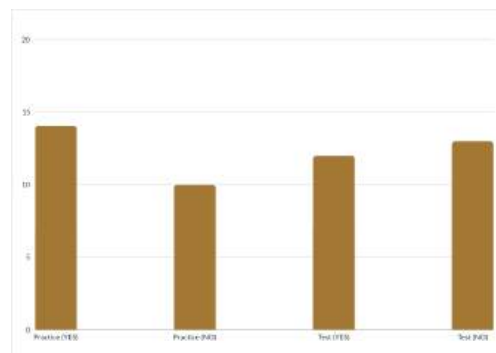


This is the average trend for the Kha character accuracies. The Y axis represents the number of participants and the X axis represents the accuracy. In the practice sessions, the majority of the participants were able to guess the correct phonetic spelling of the Kha character by hearing it pronounced. However, although the average correct answers stayed higher than the incorrect answers, the overall trend differed from this otherwise positive trend. This can be seen in the increase in incorrect answers and the decrease in correct answers. This showed that Duolingo was not able to aid these participants overall in retaining this specific character in their short term memory.

Next, the participant accuracies down each column for each character were averaged. Each of the 30 character's average accuracies for the test column was calculated as well. I then made bar graphs to compare how the average accuracy for the characters changed between the practice sessions and the cumulative tests. One again, the Y-axis represents the number of participants and the X-axis represents the accuracy. The "YES" meaning they answered correctly and "NO" meaning they answered incorrectly. The average accuracy data for week one/ trial one is as follows:



The same trend can be observed in this graph as well. Although there were more correct accuracies than inaccuracies for both the practice sessions and the test, the overall accuracy decreased while the overall inaccuracy increased. This again proves that Duolingo was not able to efficiently instill these Hindi characters in the short term members of these participants. This same trend can be observed for week/trial two:



Overall, the average accuracies decreased and average inaccuracies increased. This trend is present in both trials/weeks of this study. The presence of this trend in both weeks/trials increases the validity of the results and reduces chance. By drawing these trends back to the random sample trend for the kha character, it can be seen that they correspond closely with each other as well. These overall character accuracy trends also showed an inverse relationship between time and accuracy. In other words, as more time passed, the lower the accuracy average became. This inverse relationship is indicative of Duolingo's lack of ability to instill these Hindi characters in the short-term memories of non-native speakers. Due to the overall decrease in correct answers, Duolingo can be considered slightly detrimental to learners when aiming to improve their memorization of the Hindi language characters.

Conclusion

This research project aimed to test Duolingo's Hindi curriculum and its ability to instill Hindi characters in the short term memories of non native Hindi learners. Through utilizing the evaluative research method, this study was able to test the reliability of a singular aspect of Duolingo's Hindi curriculum, short-term memory retention. Through the findings of this study, the reliability of Duolingo's Hindi curriculum can be questioned. This study was able to present the new understanding that although Duolingo is ranked the number one language learning app on both the Google and Apple app stores, its ability to help learners remember these characters is overall low and slightly detrimental to their Hindi education. It ultimately aids in filling the gap of the lack of research on Duolingo Hindi curriculum's short term memory component and through filling this gap, guides further linguistic research and e-learning approaches for Hindi learners.

This study opens up discussion for future implications for Duolingo's Hindi curriculum. One future implication of this research is that non-native Hindi speakers might gradually look towards newer and more beneficial resources to aid in their Hindi learning journey now that they know Duolingo is not as reliable for Hindi memorization as it may seem. This is an important implication, because education plays a significant role in a linguist's language learning journey. Without the correct education, a linguist risks wasting time and learning the incorrect information ultimately resulting in drawbacks. Another future implication is that more people might become more familiar with the Hindi language as well. For non-natives speakers of the language or those completely unfamiliar with it, this study and its findings might spark curiosity and foster a connection between something they are familiar with, technology, and a potentially unfamiliar topic, Hindi. It shines light onto the complexities present in utilizing e-learning platforms to learn a language and the different perspectives to consider when deciding how to obtain the right education for it. A final implication is also a potential decrease in Duolingo user counts. Duolingo currently has over 500 million users(Piedrahíta), so the 25 participants in this study will not significantly impact these user counts; however, out of the 25 participants in this study, only 14 said that they felt as though this application aided somewhat in their learning of the language. Additionally, all 25 participants expressed that they were not interested in

continuing their Hindi education through Duolingo following this study due to various different factors.

This study also comes with a variety of limitations. A limitation surrounding my research lies in its sample size. Although a sample size of 25 was representative of the focus group of this specific study, it is not a representative sample of the entirety of the 500 million plus users on the Duolingo app. It is also not representative of the total number of Hindi users on Duolingo. Therefore, this limits the overall generalizability of my data on a global scale. Future studies could expand the sample size while implicating the same method to obtain more generalizable results and trends. A second limitation was the fact that this study did not test all the aspects of the Hindi language that are taught on Duolingo. It only tested memorization of the characters, but not speaking, writing, etc. Therefore, this study can not confidently conclude that Duolingo is overall a detrimental resource to its Hindi curriculum users. It can only express the reliability of the app in its ability to help non native Hindi users memorize the Hindi characters. The Duolingo Additionally, this study focused on short-term memory, but learning a language is a long-term process overall. With years to months of practice, language learning also relies heavily on the long-term memory retention component. That being said, future studies should investigate Duolingo's reliability in terms of long term memory retention. This could be done by conducting more trials which would mean running the study for a time period longer than two weeks. An additional limitation would be the ability to make a large scale difference through the findings of this study. My study found that Duolingo generally did not efficiently instill Hindi characters in the short-term memories of non-native Hindi speakers. However, due to Duolingo being a well established and successful company, there is an extremely low chance that curriculum changes will be made to meet the needs of the participants according to this study. A final limitation is the general presence of extraneous/confounding variables. These are factors in this research project that were not able to be controlled. These variables could be present in a variety of ways, one of which being that the participants' learning outside the lessons cannot be controlled. It was encouraged of the participants not to do any outside outside learning other than what they did in the zoom lessons, however the possibility that they furthered their learning outside this study and of the zoom sessions still exists. Another extraneous variable could be honesty. Although all participants were informed that they must be completely new to the Hindi language to participate in this study before they received the informed consent forms, chances of dishonesty exist. This dishonesty would have potential to skew the data for this study, as previously experienced Hindi learners might potentially display higher accuracies than those foreign to the language.

Overall, by discussing the limitations and future implications of this research study, it aids in guiding further research in the e-learning and linguistic fields. New Hindi learners can now have a better understanding of the Duolingo Hindi curriculum if they choose to use it in their language learning journey.

Works Cited

- "The Definite Definition of Gamification (+8 Real-World Examples)." Growth Engineering Co, www.growthengineering.co.uk/definition-of-gamification/#:~:text=%E2%80%9CGamification%20is%20the%20application%20of,applies%20them%20to%20other%20contexts. Accessed 2 Feb. 2024.
- Edmundson, Andrea. Globalized E-Learning Cultural Challenges. Google Books, books.google.com/books?hl=en&lr=&id=YXK9AQAAQBAJ&oi=fnd&pg=PR1&dq=why+does+e-learning+take+away+culture&ots=dWiYEFt0U7&sig=ODM7zU87LhmsBpWCUN8FF6-iXBE#v=onepage&q=why%20does%20e-learning%20take%20away%20culture&f=false. Accessed 22 Jan. 2024.
- Gorden, Dan Van. "Duolingo Course Review 2023: The Best and Worse Duolingo Courses." Medium, 4 Feb. 2023, medium.com/@danvangorden/duolingo-course-review-2023-the-best-and-worse-duolingo-courses-dd4121dc8461. Accessed 16 Jan. 2024.
- Jiang, Xiangying, et al. "Developing reading and listening proficiency with Duolingo: A cross-sectional analysis." Duolingo, 16 Nov. 2022, duolingo-papers.s3.amazonaws.com/reports/Duolingo_whitepaper_language_read_listen_A2_to_B1_2022.pdf. Accessed 3 Feb. 2024.
- Kentnor, Hope. "Distance Education and the Evolution of Online Learning in the United States." University of Denver Digital Commons @ DU, digitalcommons.du.edu/cgi/viewcontent.cgi?article=1026&context=law_facpub. Accessed 6 Jan. 2024.
- Kramsch, Claire. Foreign Language Research in Cross-cultural Perspective. Edited by Kees De Bot De Bot and Ralph B. Ginsberg. Google Books, books.google.com/books?hl=en&lr=&id=13RXNtwK_JoC&oi=fnd&pg=PA217&dq=how+does+learning+about+culture+help+you+learn+the+language+&ots=RweuMJnHLr&sig=q2WrLv6EKr-dRFByKqqpk3KrKV8#v=onepage&q&f=false. Accessed 5 Feb. 2024.
- Mogavi, Reza Hadi, et al. When Gamification Spoils Your Learning: A Qualitative Case Study of Gamification Misuse in a Language-Learning App, 1 June 2022. Association for Computer Machinery Library, <https://doi.org/10.1145/3491140.3528274>.
- Nicholson, Paul. A History of E-Learning. Springer Netherlands, 2007. Computers and Education. SpringerLink, https://doi.org/10.1007/978-1-4020-4914-9_1.
- Piedrahíta, Paula. "Essential Duolingo Statistics for 2024: Monthly Users, Revenue, and Time Spent." Cooljugator, 13 Nov. 2023, cooljugator.com/blog/duolingo-statistics/. Accessed 17 Feb. 2024.
- Photopoulos, Panos, et al. "Remote and In-Person Learning: Utility versus Social Experience." SN Computer Science, vol. 4, no. 2, 21 Dec. 2022. National Library of Medicine, <https://doi.org/10.1007/s42979-022-01539-6>.
- Sathyanarayanan, C. S. "Hindi and the Development of Modern Culture in India." Asian Studies, vol. XI. ProQuest,

www.proquest.com/openview/3e4f1a19f44bf0ffe53612f8f25a3fec/1?pq-origsite=gscholar&cbl=1816776. Accessed 6 Feb. 2024.

Scacchi, Marco. "Duolingo's Gamified Success: A Language Learning Triumph." *SensorTower*, Aug. 2023, sensortower.com/blog/duolingos-gamified-success-a-language-learning-triumph. Accessed 5 Feb. 2024.

Stecula, Kinga, and Radosław Wolniak. "Advantages and Disadvantages of E-Learning Innovations during COVID-19 Pandemic in Higher Education in Poland." *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 3, Sept. 2022, p. 159. National Library of Medicine, <https://doi.org/10.3390/joitmc8030159>. Accessed 3 Jan. 2024.

Suchman, Edward A. *Evaluative Research: Principles and Practice in Public Service and Social Work*. Google Books, books.google.com/books?hl=en&lr=&id=AXy6AwAAQBAJ&oi=fnd&pg=PR5&dq=evaluative+research+definition&ots=bJPn0IjmTR&sig=h_Tv6TAQXNuijkzEgLIflnFLB3c#v=onepage&q&f=false. Accessed 8 Jan. 2024.

The Impact of Artificial Intelligence on the Landscape of Law By Nicole Pelletier

Abstract

Artificial Intelligence (AI) has completely changed the way in which society functions and has improved many professions and everyday activities. However, for its benefits to society to outweigh its threats, AI cannot continue to be above the law. Instead, its potential must be bound by established parameters. This paper aims to study the novel legal risks AI poses and how these issues can be combated through enforced regulation. The evolving force of technology has begun to and will continue to have profound impacts on the landscape of law, yet legislators are delayed in their control over machine learning models. AI has unique capacities that are not quite understood, yet some of its unfamiliar capabilities are unethical and in some cases need to be defined as illegal. Further, the ambiguous nature of crimes committed by machine learning models leaves the blame asserted on the humans that develop them. Current litigation surrounding malfeasance by AI remains insubstantial and is limited to scarce lawsuits concerning copyright laws and fabricated information. Thus, it is up to the creators and users to be proactively virtuous in their operation of these systems to avoid legal repercussions. As the scope widens around legal cases of this focus, world leaders must simultaneously work to effectively navigate and create limits around Artificial Intelligence. This work investigates the varying legislative prototypes and approaches to regulating Artificial Intelligence, including the European Union's guidelines for machine learning, the United States' Blueprint for an AI Bill of Rights, and similar templates by other nations and technology organizations. This paper emphasizes the significance of the current and future generations of lawyers and policy makers' responsibility to not only regulate AI, but to grasp a full understanding of its implications on the future of the law.

Key Terms: Artificial Intelligence, Law, Legal Field, Regulation

Introduction

The field of law has existed for centuries, yet it has never ceased to adapt to novel transgressions. Alternatively, Artificial Intelligence (AI) is a relatively new innovation, but it is clear that AI is affecting every area of life, from school to work to home. Many professions are benefiting enormously from it, yet one carries the capacity to do more than profit or be damaged by AI. The politicians, lawyers, and judges who are responsible for implementing and regulating legislation are now tasked with the opportunity to control the way Artificial Intelligence is used to change and better the legal apparatus (AI Task Force, ABA). When considering the legal ramifications of AI, one must first recognize the scope of these machines' aptitudes. To elaborate, machine learning models are not granted the same rights or obligations as humans; they are not granted Locke's inalienable rights, nor are they required to act upon a moral compass. However, does this mean that these machines are above the law? Whose jurisdiction do AI machines fall under, and when this artificial technology engages in wrongdoings, who is to

pay? What so-called crimes can they potentially commit, and how many will go undiscovered? There is responsibility that must be taken, restitutions that must be accounted for, and statutes that must be altered. Artificial Intelligence is not the first new invention of its kind. Comparatively to social media, the technology's breadth and capabilities are largely unknown, but world leaders must take early action to regulate the systems before they get out of hand. Despite the many advantages of AI, the newest digital technology poses striking threats to the legal profession.

Current Law

Artificial Intelligence - specifically Generative AI - is capable of most things that most humans can do. It can create completely authentic and new content, including text, audio, video, and images. However, these machine learning systems raise concerns about how the law applies to them as well as the content they output, and who is culpable for their wrongdoings. Lawmakers have long struggled with regulating new technology in a way that does not interfere with existing laws. For example, in the early stages of social media, many attempts at regulation ran into conflicts with free speech laws (Comiter). The current laws in the United States (U.S.) and throughout much of the developed countries of the world make it relatively simple to convict humans of any given crime, whether it be tax fraud or assault. In any given case, a plea is given, attorneys present evidence for the jury to review, and the culprit is found innocent or guilty. But although AI is capable of committing the same crimes as humans, their conviction is not as straightforward. The Department of Justice (DOJ) asserts that corporate criminal liability can only be applied to the actions of a human being, and cannot pertain to the acts of an artificial entity (Tsao, Luskin & Lammers). Although machine-based AI systems are well equipped with the resources and intellect to commit legal violations, if the crimes were committed solely by the machines then it becomes impossible for the DOJ to assign blame to an individual. Supportable evidence in the court system requires criminal intent for the offense committed, and because AI systems cannot act with that intent, their actions reflect the intention of the person or people who created them (Phillip et al.). Therefore, the current legal framework in the federal government of the U.S. does not allow for the criminal prosecution of AI models (AI Task Force, ABA). So, even if the developers and users of a system had absolutely no intention for any criminal actions to be committed, if their machine learning model engages in illegal activities, the creators or companies that utilize them can be held at fault.

Nevertheless, this matter of liability should not leave companies and individuals reluctant to implement AI systems, which are substantially efficient and valuable. Instead, companies can take proactive measures to protect themselves against the legal risks posed by AI models. Moreover, companies must take steps to monitor their AI systems. Most importantly, they should obtain sufficient warranty from their data and service providers, and seek contracts that include indemnification - the security against legal liability towards the company that uses the provided AI product (Walsh). Furthermore, for the companies that are responsible for the ethical creation of machine learning models, the best way to minimize legal endangerment is to properly train the

employees who are developing them. Businesses need to have done their due diligence and testing of AI technologies before putting them into interior use or exploiting them for other company's use. Organizations must explore the potential risks or illegal activities of their machine learning systems, and once these unpredictabilities have been identified, companies can impose appropriate controls and modifications to mitigate them (Tsao, Luskin & Lammers). However, their efforts should not end there. It is equally important for companies to be proactive as it is for them to be reactive. Once the AI models are put into regular use, companies still need to run regular testing and make revisions to confirm that everything is working as intended. Because the laws surrounding Artificial Intelligence are so ambiguous and still developing, the absolute best protection against unintentional AI violation of the law is for the companies using the machines to take authority over their models to ensure they comply with the existing body of laws.

The Legal Risks of Artificial Intelligence

After analyzing the assigned blame of crimes committed by Artificial Intelligence, it is vital to discern what digital crimes it has the potential to perpetrate. However, it must be kept in mind that while AI is capable of creating these crimes, it is still debated whether they are culpable, or to blame, for these offenses (Phillip et al.). These systems lack consciousness and a sense of reason, making it impossible for them to possess the intent of committing a crime (Bjelajac). Although AI carries many of the traits of the human mind, it does not have the ability to understand good vs. evil in the ethical sense that the brain can. The scope of possible crimes that AI can commit at its free will and without instruction is vast and remarkable, yet many of which fall under the sphere of corporate crime.

The first example of these malpractices that AI is capable of executing is fraud. AI can be trained or learn to commit a variety of fraudulent activities, one of which is phishing attacks, through fake emails or social media profiles. Additionally, media manipulation, or deep fakes, are illegal abilities of machine learning (Ramluckan). Financial fraud, algorithmic trading and market manipulation, and identity or individual property theft are all capabilities of AI (Bjelajac). The more access AI has to legitimate documents, the easier it is for these machines to generate fraudulent entities based off of realistic ones. It can generate fake text, and manipulate images, both real and fake. It can analyze text to deduce the tone of the message. All of these actions give AI the ability to spread misinformation in the media or create fake news. The repercussions of fake news are catastrophic, and not only can misinformation manipulate public opinion, but it can alter the outcome of serious world events such as elections (Phillip et al.). Generative AI technologies have the resources to commit various crimes related to fraud, including defamation of character and monetary fraud. But unlike humans, who can commit these crimes with the proper planning and time, AI technologies can accomplish these deeds in merely a fraction of a second, and tracking them across the digital spectrum can prove to be almost impossible (Li).

One of the largest concerns that people have about AI is around data theft. As the internet and the companies that rely on it have evolved, user data has become a valuable asset that is often sold to third parties. Despite regulations aimed at protecting the privacy of users, data breaches are common and theft is still a serious concern. Now, the potential of AI must be added to the list of risk factors around data use and theft. AI can comb through vast amounts of information and extract specific data. It can commit various forms of data theft, including medical records, bank statements, workplace files, and cryptocurrency transactions (Comiter). Personal data is perhaps the least safe it has ever been online, with new technologies capable of malicious theft in a fraction of the time that humans are (Li). After this data is stolen, AI also can distribute the data to third parties or create accounts to impersonate the victims who have had their information stolen. Data breaches can easily result in identity thefts and other illegal acts that cause significant damage, financially or personally, to individuals.

When asked to envision a hacker, the average person would picture a person in a dark room typing away at a bright computer, face concealed. However, humans are not the only ones who have the tools to abuse systems or hack them. Machine learners are capable of gaining unauthorized access to systems, distributing malware to a large number of systems in a short amount of time, and overwhelming systems until failure. They can also guess passwords through brute force attacks (trying all possible combinations) or keystroke logging (Bjelajac). Machine learners can commit so many digital crimes completely on their own. They can be used to infect computers and networks and bring down web pages, and the data they steal can harm any number of people, from individuals to entire countries. For the first time, these digital attacks don't need a human behind them. Thus, as Artificial Intelligence models become more experienced, their hacking and cyber attacks will only become more successful and harder to track.

The possible crimes of AI also encompass attacks on infrastructures and traffic incidents. Artificial Intelligence can be programmed to ruin existing frameworks such as energy, transportation, water systems, and communications infrastructures (Bjelajac). Almost all of these categories, specifically those in large cities, are largely digitally based systems that, if shut down or compromised, would cause serious harm to the societies that rely on them. In the context of infrastructure attacks, AI has not only the ability to affect digital spaces, but also real-world environments and non-digital systems that merely rely on programmed mechanisms. Equally as disruptive as infrastructure attacks, sabotage of autonomous vehicles is a concern of AI devices. Traffic incidents can often result from real people breaking the law, such as through reckless driving, speeding, or driving while intoxicated. However, these tangible crimes are not the only ones that can cause accidents. AI technologies, when applied to automobiles, can be the largest threat to motor vehicle and road safety. Expert systems may not only be submitted to unintentional technical failures or shutdowns, but have the unfortunate aptitude to make perception errors, such as misestimation of distances or speeds of other vehicles, make wrong decisions based on analysis of other data, or malfunction during unusual situations, such as weather emergencies (Bjelajac). All of these occurrences can lead to vehicular accidents and

injury/death of innocent civilians. While the law would usually be direct and adamant in regards to the responsible party during motor vehicle accidents, in these cases where the “driver” was AI technology, who is to be held responsible remains unsettled.

Overall, the misuse of AI is inherently dangerous and in some cases criminal. Such uses of AI, including fraudulent acts, data and information theft, hacking and piracy, infrastructure attacks, and autonomous vehicle accidents, are ways in which machine learning devices can be serious threats to society (Ramluckan). Thus, these so-called crimes are potent challenges for lawmakers, and the navigation of AI within the lens of the landscape of law is moving too quickly to be effectively regulated. Because of the limited legislation on AI, it lies in the hands of the technology companies that produce machine learning and the users of it to make decisions ethically and abstain from abusing the capacities granted to them by the machines.

Existing AI Lawsuits

It is undeniable that an AI model cannot be prosecuted and sent to prison. So, attorneys are left with the challenge of figuring out how and whom to prosecute for the crimes committed by artificial entities. Novel questions have been raised about how the law applies to these machines, and some are beginning to be answered. The majority of legal proceedings surrounding AI revolve around data use, specifically in the way that generative AI harnesses pre-existing information and training to create new digital substances.

Open-source software states that when a given program is released with its source code, it allows other users to utilize it for almost any purpose, including for redistribution or modification. However, under compliance with open software regulations, users typically must give credit (called an attribution statement) to the initial author. Demonstratively, AI models without the proper training around the ethical use of open-source code fail to comply with the regulations in place and as a result, can commit plagiarism. A group of coders worked together to bring a class action lawsuit against big technology companies Microsoft, GitHub, and OpenAI, suing their illicit training of the AI tool GitHub Copilot (Walsh). The tool was trained using appreciable lines of open-source code, and it reworked the code to generate new outputs, but failed to comply with requirements. The lawsuit charged the companies with breaching software license copyright laws.

In regards to copyright laws, the most significant issue concerning generative AI materials is what copyright has the power to protect. While the U.S. Patent Office has kept unclear on whether patent laws will apply to outputs from AI systems, the U.S. Copyright Office has avowed that works wholly created by generative AI are not protectable by law because according to current laws, inventors of any given entity must be humans (Walsh). This gray area is just one example of how the current legal framework in the United States proves to be outdated when it comes to the novel challenges of AI. Accordingly, in a current case, *Thaler v. Perlmutter*, the person who invented a generative AI tool sued the U.S. Copyright Office for refusing to register a digital image that their AI tool generated (Hodson). Not unexpectedly, the District Court ruled in favor of the U.S. Copyright Office, and the registration for copyright was

denied on the basis the work was generated entirely by Artificial Intelligence, and only human created works could be registered (Hodson). Although the future remains uncertain in terms of AI copyright laws, currently the Copyright Office is steadfast in its policies to only copyright entities produced by human beings.

The legal profession must take extreme caution about the abilities and limitations of AI, and lawyers must recognize that these machines will have a serious impact on their field. Thus far, the technology has made its mark on the profession through the sanction of two lawyers due to their submission of an AI generated legal brief containing fictitious citations. The New York federal judge stated that the attorney's behavior was unethical not just because of their use of ChatGBT to draft the brief, but due to their failure to come forward upon discovering the forged quotes and citations (Mangan). This type of error made by an Artificial Intelligence model is an example of a hallucination, the tendency of models such as ChatGBT to provide inaccurate results. Although many AI models are different, most generative Artificial Intelligence prototypes have the ability to produce fabricated outputs that can only be validated or voided by the users. Concerns about these named hallucinations are important to note for all users of AI systems, but are especially relevant for lawyers intending to use findings in the court of law. To combat these concerns, some judges have announced rules for their chambers that obligate attorneys to announce their use of generative AI tools in their briefings (Burns).

These decisions depict the farsightedness that must be heeded by legal professionals to set their regulations on AI, while also working to gauge a deeper understanding of its risks, including data privacy risks, fraud, copyright infringement, and piracy. AI is impacting all parts of society, and it is up to members of the legal field to guide its regulation (Bridgesmith et al.). Thus, law schools should expand their curriculum to the crossroads of AI and the law. Students studying law need to fully grasp the changes that AI will have on their profession, and how to address these unique challenges (AI Task Force, ABA). The more emphasis that law schools put on the means in which lawyers should maneuver the impact of machine learning models, the more efficiently these machines can be navigated. It is the job of the next generation to understand and constitute Artificial Intelligence. Despite the benefits that AI has the potential to bring to the field of law, attorneys must take steps to protect themselves, their clients, and the judicial system from the threats posed to it. Extensive analysis of the technology's development and ethics, as well as its specific uses related to the realm of law, are required to establish initiatives to control the technology.

Governance and Regulation

The abilities of Artificial Intelligence are vast and unprecedented, yet AI undeniably brings about numerous consequential legal and ethical risks (Rodrigues). From its creation of fictitious outputs to committing monetary or identity fraud, it is clear that machine learning models can commit many of the same crimes as humans. They have the potential to steal personal data, disrupt infrastructures from a digital stance, and gain unauthorized access to infiltrate computer systems. Although it is evident that these tools have and can commit these

wrongdoings, legislation to prosecute AI technology and its cohorts remains tenuous. To combat these issues, governments and leaders around the world are working to accomplish effective regulations and laws around Artificial Intelligence.

Perhaps the most pioneering of these governments is not indeed a single nation but is rather an international organization. The European Union (EU) and the countries that make it up are major players on the world stage, with the means to carry profound influence over other states around the globe. The EU's relatively new regulations concerning machine learning are groundbreaking, demonstrating their international leadership and originality. Specifically, the EU has broadened its existing legal framework to accommodate the breadth of AI's new legal challenges. In 2019, the organization created guidelines for machine learning. They deemed that for AI to be trustworthy, it should be within the parameters of all existing laws and codes, be ethical, and be robust, both socially and technically (European Commission). The Union set up an agenda of seven vital requirements for reliable Artificial Intelligence.

The first measure that must be taken is ensuring that digital technology will not restrict any person's fundamental rights. As defined by the European Union, fundamental rights are essentially human rights and apply to all human beings (EU Agency for Fundamental Rights). In order to protect EU members and their nations' civilians, fundamental rights impact assessments are required to be taken both before development, as well as after and throughout the lifespan of AI tools. Additionally, the machines cannot be in full control at any time and must have human oversight. The controllers of the AI machinery must have the ultimate decision making power and the ability to override the machines (European Parliament). Next, the systems must be capable and durable, to subside the effects of cyber attacks and hacking (European Parliament). Throughout the building and finetuning of AI machinery, the developers should take all possible or apparent vulnerabilities and risks into account, not release their developments until they are accounted for. It is the responsibility of those who create the machines that simulate human intelligence to make certain of their robustness.

Thirdly, the EU advises all relevant parties in the AI community to maximize data protection and privacy in their products (European Parliament). The Union emphasizes that in no circumstances should citizens' data be exploited or used to discriminate against them. Design techniques that can be applied to ensure these safety measures include data anonymization and encryption. An additional key regulation is full transparency, to verify the unbiasedness and ethicality of models (European Parliament). Data sets should be published in their entirety and easy to find, and AI systems should make those interacting with them aware that they are contacting an artificial source that holds limited capabilities. Next, AI must be diverse and unbiased (European Parliament). Adequately, a tactical solution to the issue of bias is to expose AI systems to the entire scope of abilities and traits of the human race. One example of how bias could occur is through an inadequate data set. It is up to AI developers to ensure that their systems designs are not biased and are available and fair to all people, including those with disabilities.

Further, developers should create technology that strives for a more sustainable environment and society (European Parliament). As climate change becomes an evermore prevalent matter, new technologies should be at least partially aimed at bettering the status of the natural world. AI must be obliged to choose renewable energy systems and consider their impact on overall social and mental well-being. Finally, the legal risks of AI can be minimized by putting procedures in place to warrant full accountability for their consequences (European Parliament). Internal and external inspections of systems should be often, and reports of negative impacts should be available.

Falling in line with their overseas allies, the United States has started its own regulation of Artificial Intelligence. In October 2023, U.S. President Joe Biden signed an executive order (EO) intending to provide safeguards against AI for the public. The EO mandates developers to report outcomes of safety tests for particular AI systems that pose harm to the U.S. economy, national security, state of healthcare, or overall safety of the public (Shein). The executive branch vows to protect citizens and their privacy with this order, additionally aiming to provide further transparency and details into the ways the government is handling AI. Further, the Department of Commerce has been commissioned to authorize and guide developers to clearly label content generated artificially, as well as other AI tools and applications (Shein). Specifically, this openness serves to prevent forms of deception and fraud towards the public at the hands of AI. In summary, the Biden Administration's passage of this EO has certainly been a step closer to mandating AI, yet is not even close to a concrete and long-term solution for the problems posed by the technology.

Also encompassing the United State's legislative plan for AI is the executive branch's Blueprint for an AI Bill of Rights. The Bill is expectedly similar to the EU's guidelines and is built on many of the same fundamentals. Made up of five principles that should be followed to create the safest and best AI models, the blueprint was created to help protect the public's rights, liberties, and privacy. Firstly and most importantly, the government claims that people should be protected from unsafe or ineffective AI technology. To ensure this, systems need to undergo and report safety tests and outcomes would be classified as hazardous if systems went outside of the bounds of their intended use (The White House). Whether through the intentionally harmful design of a system or independent and unintentional growth, unsafe systems should be shut down immediately. Safe AI models should not use their ability or data to be harmful to any citizens. Next, systems should be designed equitably and no one group should face discrimination or misrepresentation. Artificial Intelligence technology should not discriminate on the basis of race, color, ethnicity, sex, religion, or any other classification protected by federal law (The White House). It is the responsibility of the developers of these automated systems to take proactive measures to promote equality, including using diverse and representative data sets when training new and existing models. Not unlike any other public service, AI should accommodate all people and have algorithmic discrimination prevention. Thirdly, data privacy should be a priority for these systems and the people whose data is being used should have control over how it is being utilized. The blueprint maintains that data protection structures should be built into all systems

and permission must be obtained for the use of personal data (The White House). Furthermore, consent requests should be understandable and clear to the public, and the decision to consent to data collection should not impact the users' availability to the system. Data collection should be limited to only what is necessary for the specific capacity of the machinery. Moreover, the users of any given system should have easy access to documentation outlining the function and roles of it, and how it impacts themselves and their peers (The White House). Plain language documentation that describes the uses and outcomes should be published simultaneously with the release of AI developments. Full transparency should not just be limited to the release of particular AI tools but should be maintained through regular updates of documentation about the machines. The last principle of regulated AI is the option of all citizens to have a human alternative in place of these systems when they fail (The White House). Users should have access to a person who can fix or explain problems that occur with the machine learning systems and this type of fallback should not pose a significant burden. The Blueprint for the AI Bill of Rights proposes many effective ways to create ethical Artificial Intelligence. However, this bill has not yet been passed into law and is therefore non-binding and largely ineffective for the legitimate regulation of the systems. If the bill does pass, the presidential administration will have to take further steps to break down each principle further into quantifiable ways to test the security of each AI model.

This paper goes into depth on two of the leading AI regulatory models (the EU and the U.S.), but it is worth mentioning that many other nations like Brazil, Canada, China, and Australia, among others, have begun to set their own limitations on AI technology (Shein). Although legislation has been thus far lacking, most world leaders are in general agreement that AI should be regulated, despite varying approaches. As previously mentioned, it is up to technology companies to regulate their own AI models and assert key principles regarding their development. Hence, companies like Microsoft have piloted computer companies' efforts to guide ethical and legal AI, laying down principles surrounding responsible machine learning models. These include fairness, reliability, safety, privacy, inclusivity, transparency, and accountability (Bjelajac). Many of the organizations, nations, or companies that have released guidelines for responsible AI have principles that corroborate with each other, pointing to the common goal of managing ethical and safe Artificial Intelligence. Most importantly, because nobody knows the exact capabilities or limits of AI, most agree it is best to err on the side of caution, and it is best to set more regulations than taking the risk of not controlling these machines enough.

The Future

This paper aims to investigate Artificial Intelligence through the legal lens. The criminal capacities of AI were examined, including its abilities to commit fraud, data and privacy theft, hacking, infrastructure damage, and traffic accidents. With these wrongdoings, the assertion of blame and legal responsibility was determined to fall on the creators and users of the AI technologies, rather than the prototypes themselves (Phillip et al.). Congruently, the limited legal

cases centered around AI have named the perpetrators of the crimes, whether it be copyright infringements or fraud, to be the companies responsible for the creation of the machine learning devices or the users of the tools. In the legal realm, AI must be understood and legally defined by attorneys, judges, and lawmakers. The future of Artificial Intelligence must be proactively controlled by not only lawmakers and governments but also by the next generation of lawyers. Law students must analyze the impacts of the newest technology on their planned profession and must work to pursue ethical AI (AI Task Force, ABA). The extent to which law schools encompass the regulations and legal mechanics of AI will shape their students' success in the field, as machine learning technology continues to grow and cases concerning its ethics and legality become evermore prevalent. Although the participation and engrossment of students are vital within the intersection of law and Artificial Intelligence, for real transformative AI development to occur, global leaders must take action. Currently, regulations surrounding the new technology are extremely insubstantial, and the way in which the law will apply to AI remains unknown. The European Union has put forward key guidelines for developers to produce ethical AI. The seven principles include human control and oversight, technological and social robustness, privacy and data protection, transparency, elimination of bias, environmental sustainability, and accountability (European Commission). Unsurprisingly, the United States soon followed suit, releasing its own Blueprint for an AI Bill of Rights with similar key principles, including fairness, privacy, safety, transparency, and human alternatives (The White House). Although still in the early stages, elected officials on the world stage are beginning to do their part regarding the regulation of the rapidly advancing technology. Other countries, such as Brazil, Canada, China, and Australia, are establishing their own regulatory démarches to combat the risks of AI. Although approaches vary in different nations based on existing policies and differing individual liberties, world leaders agree that AI must be regulated in some way before the damage becomes irreversible.

Society has many measures that must be taken and breakthroughs that must be made to achieve ethical AI. The pathway to effective regulation of these devices must involve the interdisciplinary engagement of multiple fields, namely law, government, public policy, technology, computer science, and psychology. The manner in which AI is regulated will not only shape the field of law, but all of society and the professions within it (Rodrigues). This is precisely why the exploration and development of regulations cannot merely be confined to legal professionals, but must instead stretch across and intermix many disciplines.

Works Cited

- Bjelajac, Željko, et al. "Can AI be Evil: The Criminal Capacities of ANI." *International Journal of Cognitive Research in Science, Engineering and Education*, 2023, https://www.researchgate.net/publication/376651375_Can_AI_be_Evil_The_Criminal_Capacities_of_ANI#:~:text=A%20pivotal%20dimension%20of%20ANI%27s,%2C%20and%20consequential%20decision%2Dmaking.
- "Blueprint for an AI Bill of Rights." The White House, The United States Government, 22 Nov. 2023, www.whitehouse.gov/ostp/ai-bill-of-rights/.
- Bridgesmith, Larry, and Adel Elmessiry. "The Digital Transformation of Law: Are We Prepared for Artificially Intelligent Legal Practice? ." *Akron Law Review*, vol. 54, no. 4, 2021, <https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=2546&context=akronlawreview>.
- Comiter, Marcus. "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It." Belfer Center for Science and International Affairs, Harvard Kennedy School, Aug. 2019.
- "Ethics Guidelines for Trustworthy Ai." European Commission, 8 Apr. 2019, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
- "EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation." European Parliament, 19 Sept. 2019, [www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2019\)640163](http://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2019)640163).
- "Highlight of the Issues." AI Task Force, American Bar Association, 2024, www.americanbar.org/groups/leadership/office_of_the_president/artificial-intelligence/issues/. Accessed 6 Aug. 2024.
- Hodson, Julia. "Ask the Expert: What Are Legal Issues Surrounding AI, Its Impact on the Arts?" *News at IU*, news.iu.edu, 11 July 2023, news.iu.edu/live/news/31782-ask-the-expert-what-are-legal-issues-surrounding-ai-it.
- Kelly, Phillip, et al. "Man vs Machine: Legal liability in Artificial Intelligence contracts and the challenges that can arise." *DLA Piper*, 6 Oct. 2021, <https://www.dlapiper.com/en/insights/publications/2021/10/man-vs-machine-legal-liability-artificial-intelligence-contracts>.
- Li, Victor. "How Could Generative AI Radically Transform the Legal Industry?" *ABA Journal*, 15 Feb. 2024, www.abajournal.com/web/article/how-generative-ai-could-radically-transform-the-legal-industry.
- Mangan, Dan. "Judge Sanctions Lawyers for Brief Written by A.I. with Fake Citations." *CNBC*, CNBC, 22 June 2023, www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html.

Using Machine Learning to Bet on the Spread of NFL Games By Pranav Aiyar

1. Abstract

Since its legalization, sports betting has boomed into a multi-billion dollar industry in the last five years, with betting on the NFL being the most popular. Also, in that time period, the rise of AI and more complex and accurate machine-learning models has allowed computers to predict the outcome of future events at a higher accuracy than ever before and, in some cases, at a level higher than any humans could. This research explores using machine learning to predict the spread—a bet—of NFL games to see if a computer can be accurate enough to beat sportsbooks' odds. This research will explore regressions, classifiers, and neural networks' ability to predict the spread of NFL games and turn a profit, given features about each team playing in the game. It also explores whether the number of epochs used to train the neural network model will affect its accuracy and, thus, profit. Finally, to mimic humans' tendency to only bet on games they are confident about, a confidence threshold is implemented, and accuracy and profit are recalculated.

2. Introduction

In 2018, the Supreme Court overturned a federal ban on sports betting, legalizing \$150 billion of bets Americans had previously placed illegally each year.¹ Now legal in 33 states, sports betting has become very lucrative, with betting on the NFL being the most popular. In fact, 81% of sports bettors over 18 placed at least one bet on the NFL during 2022.²

In 2021, playUSA predicted that more than \$12 billion would be bet on the NFL during the 2022 season, and last season, more than 46 million Americans placed at least one bet on the NFL.³ This boom in the popularity of betting on the NFL came after the NFL officially partnered with three sportsbooks in 2021 after years of fighting legalized sports betting.

There are many things that a bettor can wager on during an NFL game, but the point spread—called the “king of NFL betting picks”—is what this paper focuses on.⁴ The point spread is a handicap assigned to one of the teams, for betting purposes only, designed to give each team an equal chance of winning in the eyes of the bettor. For example, a sportsbook might install Team A as a 10-point favorite over Team B, meaning the sportsbook believes Team A will beat Team B by 10 points. When a bettor bets on the spread, they must bet that Team A will either “cover the spread”(win by more than 10 points) or “fail to cover”(win by less than 10 points or lose altogether). If team A wins by exactly 10 points, the spread is tied—commonly referred to as a “push”—and the bettor receives his money back. To avoid this, many point spreads end in a .5, ensuring that a bet will be won or lost.

This paper aims to use machine learning models—specifically linear and logistic regression, binning, classifiers, and neural networks—that are trained on data and results from past NFL games to predict if the spread in an NFL game will be covered or not. The research aims to develop and train a model that can be accurate enough to help a bettor make a profit if the model's predictions are followed throughout an NFL season.

3. Literature Review

In 2019, Gabriel Fialho, Aline Manhães, and João Paulo Teixeira used neural networks and a support vector machine to predict the result— a win, loss, or draw— of soccer games.⁵ Both models used 3D vectors representing each of the over 200,000 games in the Open International Soccer Database as inputs. The output was another 3D vector, with the first component representing the probability of a home win, the middle component representing the probability of a draw, and the last component representing the probability of an away win. The model’s accuracy improved as the dataset became more robust: as more games and more statistics for each game were added, the model’s accuracy improved. Special consideration was placed on teams’ performances in recent games and the effect of one team’s home-field advantage.

In 2018, Stanford students Alexandre Bucquet and Vishnu Sarukkai used ML models to predict how many points each team in an NBA game would score to bet on the point spreads of games.⁶ The models used in the paper were Random Forest as a baseline, collaborative filtering, neural networks, and LSTM to predict the spread of NBA games. The model uses stats such as Points Scored, Points Scored Against, or Total Rebounds, and more complicated metrics such as Offensive Rating and Plus/Minus. Player fatigue is also accounted for through the number of days since the last game and the distance the away team traveled to get to the game. The model gives special consideration to both teams’ recent games and their previous match-ups against each other. The LSTM model had the highest testing accuracy, and this makes sense as an LSTM takes in data sequentially and, thus, would give more weight to games that were played more recently. Collaborative fitting performed very well on the training data but struggled on the testing data, indicating obvious overfitting.

To predict the margin of victory of NFL games at a higher accuracy, Jim Warner used a Gaussian process model (3D normal distribution of a random variable, margin of victory in NFL games in this case), giving special consideration to how the game in which the stadium is played could affect the visiting team.⁷ Warner did this by gathering and using data on the difference between the temperature at which the game is played and the average temperature of the visiting team’s stadium and also considering the distance the visiting team had to travel to get to the game. If either of the data points mentioned above was large, Warner deduced that the home team’s home-field advantage could significantly impact the game’s outcome. Warner trained his model both normally and using cross-validation (using different parts of the training data each time the model was trained), but this yielded similar results to training the model normally. Even though the model was 64.36% accurate in predicting the winner of a game, it was only 50.90% when predicting the result of a bet on the spread of a game, which was lower than the accuracy needed to break even while betting on games. Warner suggested that the low accuracy could be attributed to using inaccurate features to predict the “amount” of home-field advantage or that home-field advantage does not factor into the outcome of games as much as he originally predicted. Pablo Bosch compared the accuracy of regular machine learning models to deep learning models in predicting the winner of NFL games.⁸ Bosch uniquely used a system where he considered a team’s averages from the last n games, where n varies from 3-11, in addition to considering data

from the whole season up to the game that the model is predicting. Bosch also used data from the last m mutual games the two teams played (games where both teams played the same opponent) and ranged m from 2 to 7. Bosch used a support vector machine, logistic regression, and a Random Forest Classifier for standard machine learning models. He used two neural networks— an ANN and an RNN (recurrent neural network)— and an LSTM for his deep learning models. Bosch found that the deep learning models did not outperform the logistic regression and classifiers, even after hyperparameter tuning, to try and increase the accuracy of the deep learning models. However, Bosch concluded that the LSTM was the most robust model due to it having the smallest 95% confidence interval for accuracy, indicating that with further hyperparameter tuning, an LSTM could have produced the highest accuracy.

4. Methods

4.1 Data

The data used for research is a combination of 2 datasets: one containing the results of NFL games from the past 4 NFL seasons (2019-2022)⁹ and a dataset of FiveThirtyEight’s NFL Elo ratings.¹⁰ Both datasets were split into a training set (2019-2021 data) and a testing set from 2022. For the purpose of this research, playoff games were removed from the dataset even though they were included originally. Additional columns— including both teams’ wins until that point during the season and both teams’ total wins from the previous season— were added to the dataset using Excel and Python functions. All in all, there are 783 individual games used for training and 270 used for testing. There are 48 columns in the final dataset, but only 15 numeric columns are used to predict the value of the ‘spread covered’ column, a boolean stored in numeric form. The numerical columns that were passed into all of the models used in this project are below (Table 1 & Table 2).

Variable Name	Variable Explanation
schedule week -	The week in which the game is happening
spread favorite -	The spread set by the bookmaker for the game.
over under line -	The over/under line (total points scored in the game) set by the sportsbook for the game.
fav W week & under W-week	These variables store each team (favorite and underdog’s) wins on the season until the week of the game being predicted. These values are much more useful when predicting games in later weeks.

The variables in the table below (Table 2) are taken from FiveThirtyEight.com’s Elo Forecasting system. The methodologies for calculating the value of each variable for each game can be found on the website.¹¹ This system is not being used for the 2023 season (as of October 2023), so the values of each variable have to be calculated as outlined on the website.¹¹

Variable Name	Variable Explanation
home/away qb elo pre	The quarterback-adjusted Elo Ratings for each team prior to the game.
elo prob home/away & qb elo prob home/away	These variables calculate the probability that each team will win the game based on their standard and qb-adjusted Elo. The probability is calculated by taking: $P (team) = \frac{1}{10^{\frac{Elo_Diff}{400}} + 1} \quad (1)$
home qb/away qb value pre	The value assigned to each team's quarterback is a rolling value that adjusts based on the QB's previous value and his stats and quarterback rating in the previous game.
home qb/away qb ad-justment	The quarterback Elo adjustment('qb adj') is applied before each game by comparing the starting QB's rolling value('qb value pre') rating with the team's rolling rating and multiplying by 3.3.

A point of note is the lack of the 'home elo pre' and 'away elo pre' columns in the features used to train the model. Because the variables are so dependent on each other (qb elo is just Elo adjusted by a certain number of points based on the calculated strength of each team's quarterback), using both sets of ratings only brought down the accuracy of every model. 'home/away qb elo pre' were used instead of 'home elo pre' and 'away elo pre' because it resulted in higher prediction accuracy when used for most of the models.

4.2 Methodology

1. *Linear Regression*: The first model I used was a simple linear regression to visualize what trends, if any, existed between certain independent variables and the dependent variable. For linear regression, where both variables need to be quantitative in nature, I used the 'spread covered amount' variable as my dependent variable instead of the boolean 'spread covered' as is done in later models. Through experimentation with single linear regression, the variables that had any correlation to the 'spread covered amount' were revealed; allowing me to narrow down the features that would be used later to train the neural network. Finally, after splitting the dataset into training and testing sets (2019-2021 for training and 2022 for testing), a multiple linear regression model using the features listed above outputted the predictions of the linear regression. The linear regression gave a quantitative prediction of how much the spread in a game would/would not be covered by. Using a for loop and if statement, the model's predictions can be turned into binary values for 'spread covered,' which could then be compared to the actual 'spread covered' column to determine the model's accuracy.

2. *Logistic Regression and Classifiers*: Using the same process and features as the linear regression model, a multiple logistic regression was run to see if there would be any differences in

accuracy compared to the linear regression. Using the same features, nine different classifier models were also tested. Hyperparameter tuning was used to maximize the accuracy of the classifiers and minimize the loss/error.

The classifiers used were K-Nearest Neighbors(KNN),¹² Support Vector Classifier(SVC),¹³ Gaussian Process,¹⁴ Decision Tree Classifier,¹⁵ Random Forrest,¹⁶ MLPClassifier,¹⁷ AdaBoost,¹⁸ GaussianNB,¹⁹ and Quadratic Discriminant Analysis.²⁰

3. *Neural Network* : The final model used in this research is an ANN(Artificial Neural Network). The neural network used in this research uses the relu activation function on the input and hidden layers before using a sigmoid function on the output layer to create a binary classification. The model outputs a value between 0 and 1, where any value below 0.5 predicts that the spread will not be covered, while a prediction over 0.5 predicts that the spread will be covered for a particular game. A value closer to 1 rather than 0.5 indicates that the model is more confident that the spread will be covered in a particular game, and a value closer to 0 than 0.5 indicates that the model strongly believes that the spread will not be covered in that game.

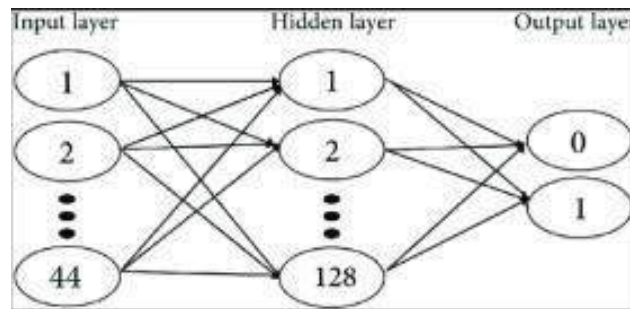
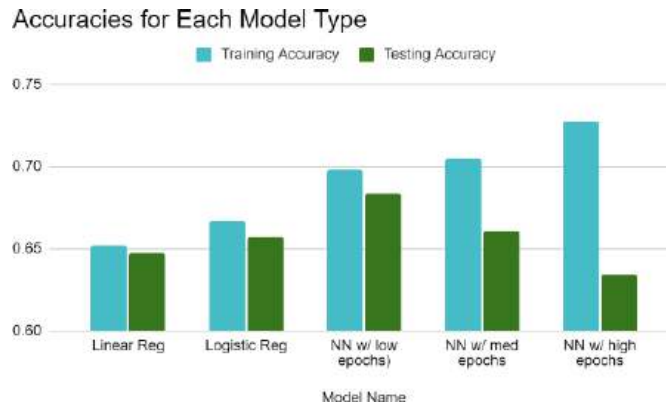


Figure 1: Example of a neural network that uses a squashing function to output a binary value of 0 or 1.

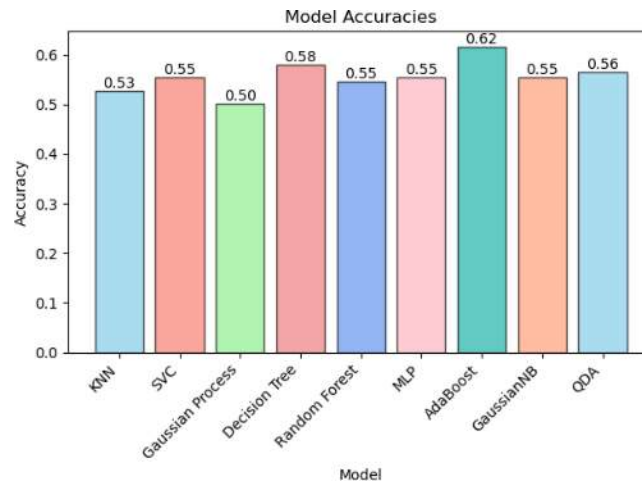
5. Results



(a) Figure 2: Graph of training and testing accuracies for each model.

Model Name	Total Profit	ROI
Linear Reg.	\$704.24	23.54%
Logistic Reg	\$759.64	25.39%
NN w/ low epochs	\$914.44	30.56%
NN w/ med. epochs	\$783.63	26.19%
NN w/ high epochs	\$632.84	21.15%

(b) Table 3: The total profit and the ROI each model would have made had it been implemented in the 2022 season. Total profit is calculated by taking $(10 * test\ acc * -11 * (1 - test\ acc)) * 272$, and ROI is calculated by doing $(test\ acc / (272 * 11))$.



(c) Figure 3: Graph of the testing accuracy of each classifier used.

The money-based calculations in this research are based on the assumption that a bettor will place \$11 on the spread of a game, with a potential winning of \$21 for a correct bet (profit of \$10). The -110 odds on each bet a bettor makes is standard for spread betting and is how sports- books make money.²¹ The number **272** is the number of regular season games played in the 2022 season.

For linear regression, the accuracy score was computed by comparing the sign of the model's prediction of how much the spread in a game will be covered by (can be a positive or negative value depending on if the model predicts the spread will be covered or not, respectively) with the sign of the 'spread covered amount' column in the dataset. If both had the same sign, the model's prediction was marked as correct. Also, for linear regression, the testing accuracy is from the model being run on only the testing data, while for both linear and logistic regression, the training accuracy is simply the model's accuracy when running on the training dataset. Using linear regression, the relationship between independent variables such as 'spread favorite' and 'over under line' (both set by sportsbooks before games) and the dependent variable of 'spread covered amount' was revealed to be positive. 'spread covered amount' vs 'spread favorite' had an $r = 0.233$, while 'spread covered amount' vs 'over under line' had $r = 0.109$ when

run on the training data. Including ‘spread favorite’ and ‘over under line’ as features in my neural networks led to positive jumps in prediction accuracy for all three epoch values I tested. The relationship between ‘spread covered amount’ vs. ‘total win diff’ (the difference between the teams’ total # of wins from the start of the dataset up until the week of the desired prediction) had an $r = -0.029$, and using the teams’ “total wins” as a feature for the neural networks led to lower accuracy.

The neural network I built was trained with three different numbers of epochs, all of which resulted in different accuracies for both the training and testing data. Training the NN with a small number of epochs resulted in the lowest training but highest testing accuracy. This trend continues, as increasing the number of epochs increases the training accuracy but decreases the testing accuracy.

Model Type	Testing Accuracy	# of Confident Preds.
NN w/ low epochs	0.727	133
NN w/ med. epochs	0.713	143
NN w/ high epochs	0.659	205

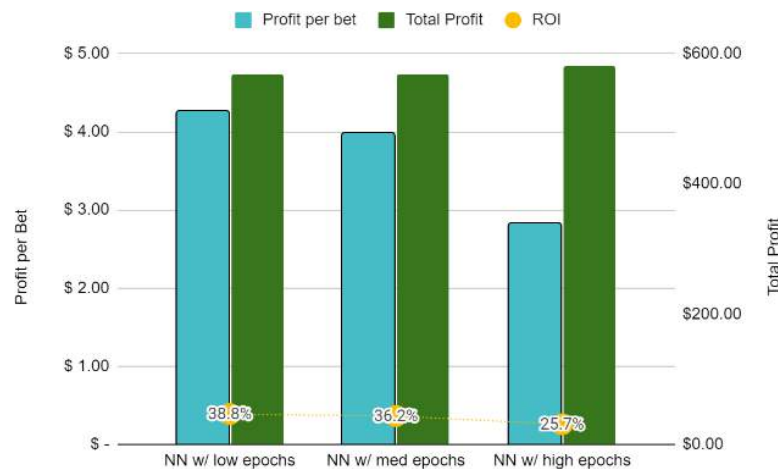


Figure 3: Table 4 & Figure 4: The table and graph above display output data from the neural networks when a confidence threshold was implemented. The thresholds used for the data in the figures are 0.35 and 0.65. This means that the testing accuracy for the NN with the lowest number of epochs was 0.7273 for predictions that were greater than 0.65 and less than 0.35. The “# of Confident Preds” column displays the number of games in which the model’s prediction was greater than 0.65 and less than 0.35. 0.65 and 0.35 were chosen as “middle-ground” confidence thresholds to offer a balance between total profit(maximized by having no confidence thresholds) and ROI(maximized by having very high confidence thresholds). The same formulas(as Figure 5) are then applied to calculate the Total Profit and ROI given that bets are only placed on “confident predictions”. The “Profit per bet” column is calculated by taking $testing\ acc/\#of\ Confident\ Preds$.

6. Discussion

6.1 Discussion

The results showcased by this research follow the trends showcased by past literature but with a significant exception. Both Warner⁷ and Bosch et al.⁸ found success with classifiers, specifically Gaussian and Random Forest. However, that was not the case with this dataset. This could be due to inaccurate hyperparameter tuning, the dataset's smaller size, or the use of different features.

Consistent with Fialho⁵ et al.'s findings, the accuracy of every model increased as input features were added. The most significant increase occurred when the Elo and the "NFL Scores and Betting Data" dataset from Kaggle were combined. This makes sense as combining the datasets was the most significant column expansion done during the research and also because the Elo ratings from 538 are very robust calculations.

Using so much quarterback data in the FiveThirtyEight Elo scores is very interesting, as all eight features from the "NFL Elo Ratings" dataset (from FiveThirtyEight) specifically account for the quarterback's performance. As described in the data section, the use of the 'qbelo home pre' and 'qbelo away pre' columns instead of the 'home elo pre' and 'away elo pre' columns resulted in a higher prediction accuracy, which goes to show just how much of a quarterback-driven league the NFL has become.

Bill Holst, the creator of the "NFL ELO Ratings" dataset, found that it was 64% accurate in predicting the winners of games and 54% accurate against the spread. Combining the Elo data with team records and the "NFL scores and betting data" significantly increased the accuracy of the models used for this research.

Experimenting with different numbers of epochs for the ANN is absent in previous literature and produced some interesting results. The trend observed in training accuracy is straightforward: as the model is trained for more epochs, its accuracy increases. However, overfitting is a problem when using a high number of epochs, as shown by the negative correlation between testing accuracy and the number of epochs.

Similar to Warner's⁷ use of confidence level, which he claims "lends itself naturally to a betting scenario where one looks to balance the risk of placing a bet on a game with their relative certainty in its outcome," with a Gaussian process classifier, a confidence threshold is used with an ANN to maximize ROI (return on investment). Betting on every game, regardless of the model's confidence, maximized total profit because the slightly lower accuracy is made up for by having 272 games to bet on, as seen by the \$914.44 profit generated by the ANN trained with the lowest number of epochs. However, using a confidence threshold with the ANN trained on the lowest number of epochs maximizes ROI, while betting based on the ANN trained on the middle number of epochs would be my preferred betting strategy, as it offers a higher total profit while also having a relatively high ROI.

6.2 Limitations and Further Work

The main limitations of this research lie in the limitations of the dataset, namely, a lack of

in-game statistics and a lack of emphasis on non-QB positions. The features that were taken from the “NFL ELO Ratings” dataset all have a focus on the quarterback, and even though team performance is incorporated into the Elo ratings and the QB adjustments, having additional data that represents a team by the strength of their skill players (running backs/wide receivers/tight ends) or defense could improve the accuracy of the models even further. This can be explored in future research, using a formula similar to the one used to compute “QB Value” to find how strong teams are at other positions/position groups. However, this does not mean QB-centric data should be removed— as it led to a higher accuracy than was achieved by Warner⁷ and Bosch et al.,⁸ both of whom used data that was focused on different factors and stats that could affect the outcome of a game— and instead, adding additional data for other positions should be experimented with.

Further exploration should also be done in changing the size of the dataset. I used data from 2019-2022 and trained my model on three years worth of games, while Warner⁷ and Bosch et al.⁸ both used eight years of data. The higher accuracy achieved in my experiments could indicate an advantage to using fewer years of data. Teams change vastly from year to year in the NFL, and game results and stats from 5-8 years in the past could be misleading and ultimately hurt prediction accuracy. The loss of accuracy when using ‘total wins’-in my research also points to a likely decrease in accuracy when heavy weight is given to the data/records of past years. However, further research would need to be done to prove this conclusively. A dataset that uses somewhere between 3 and 8 years could also be used to analyze its effect on accuracy. In addition, specific stats from teams’ previous games could be added to the dataset to improve model accuracy. The specific statistics and the games they are taken from are up to the researcher, but adding statistics such as average turnover differential and average TOP (time of possession) in the last n games (the value of n can be decided by the researcher to maximize accuracy) could be used in addition to the already used data to improve model accuracy further, and thus increase profit.

Finally, future researchers could experiment with an LSTM model to see if it gives a higher accuracy than a neural network. Both Bucquete and Sarukkai⁶ and Bosch et al.⁸ found that the deep learning model with the best accuracy was an LSTM model. A plausible explanation for why an LSTM model would have a higher accuracy is because an LSTM gives more weight to games that happened more recently when making predictions, which an ANN does not do. However, this explanation must be experimentally substantiated in future research to be proven true.

7. Conclusion

As AI & machine learning continues to advance, sports betting will quickly become a field where using machine learning to inform or even make bets will be commonplace. Especially in the NFL, where new, previously untrackable data points appear each season, AI models will become even more robust as they have more data and features to work with. As shown by the models in this research having a higher accuracy than similar models deployed in research, as well as every model turning a profit, AI in NFL betting could cause sportsbooks to

change the -110 odds currently used for most bets on the spread, necessitating an even higher accuracy to turn a profit. Just as it has in so many other areas, when using AI or machine learning models becomes commonplace, the landscape of sports betting will be changed forever.

8. Acknowledgements

Thank you for the guidance of Christian Yongwhan Lim from Columbia University in the development of this research paper.

Works Cited

- ¹ The New York Times. Supreme court ruling favors sports betting.
<https://www.nytimes.com/2018/05/14/us/politics/supreme-court-sports-betting-new-jersey.html>, 2018.
- ² A. Walsh. Which is the sport americans bet on most?
<https://www.rotowire.com/article/us-sports-betting-news-which-is-the-sport-americans-bet-on-most-71608>, [2023].
- ³ Benzinga. Sports betting check-ins jump 77% for opening night: Here's how many people could bet on 2022 nfl season. <https://t.ly/w5JXS>, [2022].
- ⁴ USA Today. Nfl picks & predictions.
<https://www.usatoday.com/betting/sports/nfl-picks-predictions-best-bets/#Spreads>, [2023].
- ⁵ J. P. Teixeira G. Fialho, A. Manhães. Predicting sports results with artificial intelligence – a proposal framework for soccer games. *Procedia Computer Science*, 2019.
- ⁶ V. Sarukkai A. Bucquet. The bank is open: Ai in sports gambling. *Stanford*, 2018.
- ⁷ J. Warner. Predicting margin of victory in nfl games: Machine learning vs. the las vegas line. *ResearchGate*, 2010.
- ⁸ P. Bosch. Predicting the winner of nfl-games using machine and deep learning. *Vrije universiteit Amsterdam*, 2018.
- ⁹ T. Crabtree. Nfl scores and betting data.
<https://www.kaggle.com/datasets/tobycrabtree/nfl-scores-and-betting-data>, 2023.
- ¹⁰ D. Fernandez. Nfl elo ratings. <https://www.kaggle.com/datasets/dtrade84/nfl-elo-ratings>, 2022.
- ¹¹ FiveThirtyEight. How our nfl predictions work.
<https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/>, [2023].
- ¹² scikit learn. Kneighborsclassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>, Accessed on: 10/15/23.
- ¹³ scikitlearn. Svc. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>, Accessed on: 10/15/23.
- ¹⁴ scikit learn. Gaussian process. https://scikit-learn.org/stable/modules/gaussian_process.html, Accessed on: 10/15/23.
- ¹⁵ scikit learn. Decisiontreeclassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>, Accessed on: 10/15/23.
- ¹⁶ scikit learn. Randomforestclassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>, Accessed on: 10/15/23.
- ¹⁷ scikit learn. Mlpclassifier. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html, Accessed on: 10/15/23.
- ¹⁸ scikit learn. Adaboostclassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>, Accessed on: 10/15/23.
- ¹⁹ Scikit Learn. Gaussiannb. https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html, Accessed on: 10/15/23.
- ²⁰ Scikit Learn. Qda. https://scikit-learn.org/stable/modules/lda_qda.html, Accessed on: 10/15/23.
- ²¹ M. De Saro and Pempus B. What do the sports betting odds mean.
<https://www.forbes.com/betting/sports-betting/what-do-sports-betting-odds-mean/>, [2023]

The Crucial Role of Eicosanoids in Liver Disease By Amber Wang

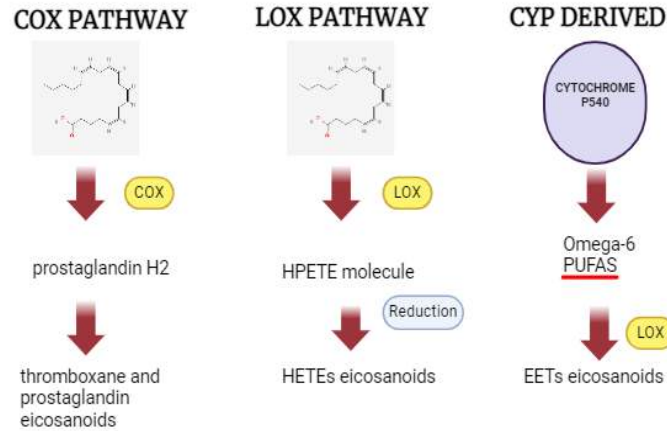
Abstract

Liver disease is a significant global health issue, leading to millions of deaths worldwide (Devarbhavi H et. al). Its complex pathology has become a major area of research. Inflammation is crucial in the development and progression of liver disease. Eicosanoids are lipid based signaling molecules involved in various cellular processes: notably hepatic inflammation. They are present in both the acute and chronic phases of liver disease. Their various functions of liver regeneration, anti-inflammatory and inflammatory resolution has made these molecules a critical area of investigation for scientists. The review aims to summarize the role of eicosanoids in liver disease: as biomarkers and potential therapeutic targets.

Biochemistry of eicosanoid pathways

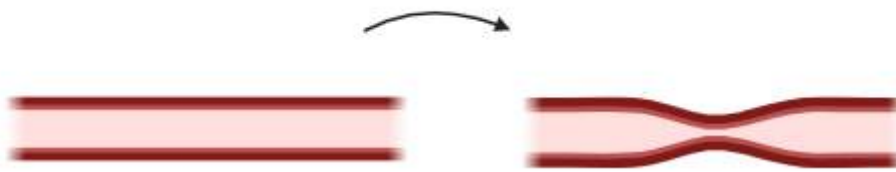
Eicosanoids are synthesized through COX, LOX, and CYP pathways. Arachidonic Acid (AA) is a key component of these pathways, as it is the precursor for eicosanoids (Shoieb et. al). In the COX pathway, arachidonic acid is converted into prostaglandin H₂ by COX enzymes. The unstable intermediate is then metabolized to form prostaglandins and thromboxane eicosanoids (Sheppe AEF et. al). LOX enzymes in the LOX pathway metabolize arachidonic acid into HPETEs, which are then reduced to HETEs. Finally, through a CYP-enzyme mediated reaction, this omega-6 polyunsaturated fat (PUFA) can produce EETs, which are involved in regulating vascular tone and inflammation (Shoieb et. al). These various classes of eicosanoids all play independent roles in the development of liver disease. EETs and Lipoxins produced via LOX enzymes are involved in anti-inflammatory functions, inversely correlated with nonalcoholic fatty liver disease (NAFLD) severity, which brings their potential for therapeutic mechanisms into question (Maciejewska et. al). HETES have major proinflammatory functions, with concentrations increasing as the severity of liver disease progresses (Maciejewska et. al).

3 Main Eicosanoid Pathways:



Eicosanoids and liver disease

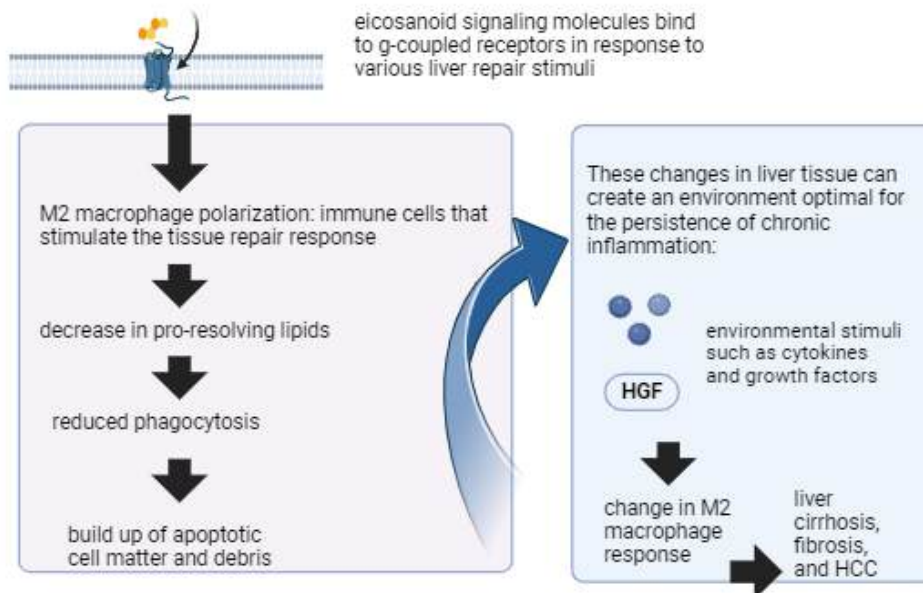
Eicosanoids influence liver disease progression in both chronic and acute stages. Acute liver damage can be caused by a variety of factors such as toxin/chemical exposure and can occur in individuals with no history of liver problems (Shoieb et. al). Acute liver disease induces proinflammatory eicosanoids produced by hepatocytes to signal the mitogenic and cell proliferation response. Liver macrophages will aid in this response. Then, oxylipins, bioactive molecules with anti-inflammatory properties, are synthesized to mediate the resolution of inflammation. These compounds counteract pro-inflammatory eicosanoids, thereby restoring hepatic homeostasis (Alba MM et. al). Additionally, thromboxanes can stimulate micro vasoconstriction, impacting blood flow and pressure in the liver and maintaining homeostasis by preventing blood loss (Alba MM et. al). This can be depicted in the image below:



Vasoconstriction increases intravascular pressure, reducing blood flow to damaged liver tissue and minimizing blood loss. This mechanism redirects oxygenated blood to healthier areas of the liver, promoting hepatocyte regeneration (Lin X et. al). By stabilizing blood flow, vasoconstriction aids in creating an optimal environment for acute liver repair (Lin X et. al). Conversely, chronic inflammation is the result of the persisted presence of pro-inflammatory eicosanoids. Eicosanoids will induce M2 macrophage polarization, attempting to produce tissue repair and anti-inflammatory responses (Alba MM et. al). However, this causes a decrease in pro-resolving lipids, which are responsible for the resolution of inflammation, promoting liver

tissue homeostasis. A reduction in these lipids reduces phagocytosis, a necessary ability of macrophages in clearing apoptotic cells and debris (Alba MM et. al). From chronic inflammation, M2 macrophages can change their responses due to changes in their environment, often making the liver susceptible to the development of fibrosis, cirrhosis, and eventually hepatocellular carcinoma (Alba MM et. al).

Progression of chronic liver disease to cirrhosis, fibrosis, and HCC:



Eicosanoids as biomarkers

Liver biopsy is a dangerous and invasive procedure for detecting liver disease. Additionally, the surgery only covers 1/50,000 of the organ (Maciejewska et. al). Therefore, non-invasive but accurate methods should be investigated to overcome the limitations of liver biopsy. The concentrations of eicosanoids vary depending on the stage of NAFLD and NASH development, making them sufficient biomarkers for liver disease. Could the use of eicosanoid concentrations become a primary tool for the detection of liver disease? Because liver diseases like NAFLD affect the whole body, not just the liver, eicosanoid samples in serum blood can be compared to eicosanoid concentrations in the liver to detect inflammation (Maciejewska et. al). Maciejewska and colleagues found strong correlations between serum and liver concentrations of 9-HODE and 13-HODE, both of which were extremely elevated, suggesting that they are good predictors of NASH and fibrosis. Concentrations of 9-HODE in plasma also differed between patients in the 1st and 2nd steatosis stages of NAFLD. The data also showed a decreased concentration, mostly in serum, of resolvin E1, which has anti-inflammatory properties. HETE eicosanoids also show promising results, although their correlation with disease progression was weaker. All HETES increased in concentration in both serum and liver samples during NAFLD progression, especially during NASH development as well (Maciejewska et. al).

The researchers also found that NAFLD progression is closely correlated to oxidative stress levels. PUFAS such as Arachidonic acid (AA), linoleic acid (LA), eicosapentaenic acid (EPA), and docosahexaenoic acid (DHA) are susceptible to oxidative damage from their multiple double bonds. Oxidative stress induces lipid peroxidation in PUFAS due to the imbalance of ROS, which can directly oxidize PUFAS into eicosanoids. Resulting eicosanoids activate further inflammation, starting an ongoing cycle that can eventually progress to severe diseases including HCC, cirrhosis, or NASH (Maciejewska et. al). These findings suggest that managing oxidative stress levels can be revolutionary in treating inflammation and liver disease.

A previous study implemented by Maciejewska and colleagues also highlighted the significant correlation between lifestyle adjustments and eicosanoid concentrations. A decrease in pro-inflammatory eicosanoids was seen in patients who followed a successful six-month dietary plan. Additionally, patients who lost greater than 7% of their body weight experienced improvements in NAFLD progression and significant decrease in pro-inflammatory eicosanoid concentrations as well (Maciejewska et. al).

Conclusion

The complex pathology of liver disease is ultimately shaped by these lipid based signaling molecules. Their significant relationships to the physiological symptoms of liver disease make eicosanoids significant biomarkers and effective therapeutic targets (notably the anti-inflammatory pathways). Although much is known about the pathways involved in the synthesis and functioning of eicosanoids, there is much improvement to be made. Components of anti-inflammatory and pro-resolving eicosanoids like resolvin and protectin can be researched for liver disease therapy. Because their concentrations decrease as liver disease progresses¹, the pathways in which they function should be further analyzed to enhance their effects when treating liver disease. Although eicosanoids are known to be involved in liver regeneration after chronic liver disease, their specific roles under these conditions are not completely understood, providing opportunities for further research.

Works Cited

- Alba, Mario M., et al. "Eicosanoids and Other Oxylipins in Liver Injury, Inflammation and Liver Cancer Development." *Frontiers in Physiology*, vol. 14, Frontiers Media, Feb. 2023, <https://doi.org/10.3389/fphys.2023.1098467>.
- Devarbhavi, Harshad, et al. "Global Burden of Liver Disease: 2023 Update." *Journal of Hepatology*, vol. 79, no. 2, Mar. 2023, <https://doi.org/10.1016/j.jhep.2023.03.017>.
- Lin, Xiaoqi, et al. "Editorial: Acute Liver Injury and Repair: Cellular and Molecular Mechanisms." *Frontiers in Cell and Developmental Biology*, vol. 12, Frontiers Media, June 2024, <https://doi.org/10.3389/fcell.2024.1439921>.
- Maciejewska, Dominika, et al. "Eicosanoids in Nonalcoholic Fatty Liver Disease (NAFLD) Progression. Do Serum Eicosanoids Profile Correspond with Liver Eicosanoids Content during NAFLD Development and Progression?" *Molecules*, vol. 25, no. 9, Apr. 2020, p. 2026, <https://doi.org/10.3390/molecules25092026>. Accessed 14 May 2021.
- Sheppe, Austin E. F., and Mariola J. Edelman. "Roles of Eicosanoids in Regulating Inflammation and Neutrophil Migration as an Innate Host Response to Bacterial Infections." *Infection and Immunity*, edited by Karen M. Ottemann, vol. 89, no. 8, July 2021, <https://doi.org/10.1128/iai.00095-21>.
- Shoieb, Sherif M., et al. "Cytochrome P450-Derived Eicosanoids and Inflammation in Liver Diseases." *Prostaglandins & Other Lipid Mediators*, vol. 147, Elsevier BV, Apr. 2020, pp. 106400–0, <https://doi.org/10.1016/j.prostaglandins.2019.106400>.
- All biological diagrams created with Bio Render: <https://www.biorender.com/>

Japanese Imperialism Through the Lens of Post-Colonial Theory

By Diying (Eva) Jiang

Abstract

Colonialism reached its height in the 20th century, not just from the western empires, but also in the eastern hemisphere, where Japan established its colonial power in Asia. A century later, the post-colonial effects from Japan is still present in the modern political dynamics. With the abundance of research that navigated the post-colonial effects and post-colonial theories for the western empires, this research aims to understand the post-colonial presence of the Japanese Empire through the lens of the post-colonial theories designed for western empires.

Introduction

Since the end of World War II(WWII), the international order has been reshaped by imperialistic history. How imperialization molded the international world remains clearly pronounced in different countries' interactions today, including how the United States and many other western former empires still establish their statues labeled “modernized”, “liberal” and “advanced”. As the postcolonialism theory connotes, the western empires have colonized eastern countries, and the legacy shapes how many of the colonial subjects and the empires themselves still view the former empires as superior, with the former empires also presuming a more advanced and liberal position even today. This bias pervades in numerous levels of the societal culture. Even in daily conversation, how certain accents are depicted as elegant and others as awkward is an example of the lasting inequality in the relationship in the post-colonization world.

Under such a layout, postcolonialism theory by Edward Said, Homi Bhabha, and many other scholars challenges such behaviors and aims to reveal the underlying pattern of how empires and colonized subjects interact in a twisted narrative, combatting the inequality amongst the ancient and modern representation of cultures.

In the field of international relation studies, the theory of postcolonialism primarily concentrates on the relationship between western empires and eastern colonies in modern times. As much as postcolonialism supposedly investigates the relationship between empires and colonized subjects, the non-western empires don't seem to draw as much attention. The Japanese Empire seemed to be considered less when evaluating the impact of imperialization in the modern world. However, is it fair to neglect the presence, or the impact of non-western empires?

I argue that the Japanese Empire's influence should not be overlooked. The acts of Japan's imperialization created a drastic impact on how the east Asian countries viewed each other, also being obvious in the political and societal dialogues today in the international society. The Japanese empire was one which lasted from 1868-1945, spreading its colony across the vast majority of Asia, leaving an unavoidable mark in Asian and world history. Different from the commonly noted western empires in the postcolonial theory, the Japanese empire was not conducted in the pronounced ambition to rule the world, rather, Japan aimed mainly at neighboring Asian countries as targets of their expanding empire. Eventually, the Japanese

developed the representation of their empire as effort to create an "Eastern Asia Co-Prosperty Sphere" (Grajdanzev). Separating from the western colonies for the similarity in ethnicity, geographical location, and cultural context between the colony and the colonized. In the process of Japan's imperialization, the acts of building comfort stations, massacre, and many other acts of violence and brutality remain a central topic of dispute amongst Asian countries. Attitude from both the citizens and the government officials of the countries that were once Japanese colonies bring forth new perspectives of how the Japanese empire, an example of a non-western empire, positions itself in the current international relations system.

Undoubtedly, the communication between Japan, China, and South Korea, does show a level of similarity to the analyzed outcomes in the traditional postcolonial works, mostly in terms of the ambivalence, altered narrative and the barrier of expression. However, the Japanese empire showed its difference with a lower emphasis on the abrogation of indigenous culture, and slight difference in the distinction between "self" and "other" in its colonial process.

As postcolonialism investigates the relationship between western empires and eastern colonized subjects, this research aims to examine the perspective of Japan's imperialism on the other Asian countries as an example of non-western Empire; specifically using China (The People's Republic of China) and South Korea (The Republic of Korea) as an example. Delving into how the empire shreds its impact in the intercommunication between three of the extremely powerful actors in Asia, all taking a major role on the global stage. This paper will evaluate the application of traditional postcolonialism works from Edward Said, Homi Bhabha, Gayatri Spivak and Bill Ashcroft, as four representative examples, on the Japanese empire's legacy. How political demographics have been influenced in the past, current and future East Asia will also be addressed.

Post-Colonial Theory

Colonization first started in the 1880s, soon becoming common amongst many of the powerful countries who aimed to expand their control to other lands. In the time of a foreign power's rulership over a country, the colonized country's culture, legislature process and values, will endure drastic changes through the "collisions, associations, and intermarriages" between the colony and the colonized. Most commonly, the empires will mark themselves as more advanced or superior to their colony, and propose that imperialization will bring "modernization" (Boehmer). With modernization further defined as the process of bringing the colony away from a "primitive" stage, and attaining a "social existence that is significantly different to all past forms of human experience" (Shilliam).

In nature, post colonialism provides a perspective to evaluate the interaction between empires and colonial subject, including how the countries would perceive each other during and after colonization, as well as how individuals' identities are altered, and many more topics. Under colonization, the indigenous population is narrated by the empire in ways that favor their propaganda and imperial rule, starting the process in which the local population is made into caricatures that disregard the complexity of their identity, beliefs, and culture (Krueger).

Post-Colonialism is a theory first created in Edward Said's "Orientalism", which evaluates the inequality and misunderstanding between western empires and their eastern colonies. Interpreting scenarios where western cultures often face challenges in understanding the foreign eastern cultures, they hence classify the eastern culture with more brutality, combined with depicting them as "mysterious" and "magical". Throughout the imperial process, the empires would attempt to convert the eastern culture in the name of "education" for "civility", molding the eastern cultures in a way so that it becomes similar with western ideals.

Based on Edward Said's work, Homi K Bhabha made the additional explanation of the existence of "ambivalence" in the intricate relationship between the colonizing power and the colonized subjects, being a connection that lies between admiration and opposition, often also noted as the "mix of attraction and repulsion" (Nasrullah Mambrol). This contribution replaced the oversimplified view of the attitude that colonized subject, and colonies hold towards each other. Even further, Bhabha developed into the concept of mimicry and hybridity (Marotta). Indicating how the colonized is unable to erase the colonized history and its influence on their country, commonly resulting in the mimicking of the colonizing culture's characteristics. This could include an increased similarity with the colonizer's values, clothing, food culture, and others.

Gayatri Chakravorty Spivak contributed to the understanding of how identities could restrain individuals from expressing themselves; understanding why the subaltern cannot speak. Spivak was able to present the term "Epistemic Violence", where the colonizer suppresses the people's ability to express themselves by representing them in an inferior position, while masking the control with the name of civility, religious redemption, and more. Such impacts show even more severely amongst groups that were treated as inferior even in the uncolonized past. Women, become a fragile target, under the social norms that demote women's statuses, women's authentic expression of their experience is also muted. Furthermore, language is also interpreted as a key lens for expression. In her work, Spivak elaborated on how many of the colonized population face the obstacle of only being able to speak through the language of the colonizing country, hence only being able to articulate their ideas with terms that the colonizers provided. The language barrier reinforces the challenge that the oppressed population face (Maggio). As Said discusses post imperialization with the colonized countries remaining silent, Spivak played an instrumental role in filling in the missing perspective of the reason that the imperialized countries remain silent.

Bill Ashcroft then built even further upon and contributed an investigation into the empire's abrogation and appropriation of the colonized culture's context. Illustrating of the term "abrogation", defined as "the refusal of the categories of the imperial culture, its aesthetic, its illusory standard of normative of 'correct' usage and its assumption of a traditional or 'fixed' meaning 'inscribed' in the words", and appropriation as "the process by which the [imperial] language is made to bear the burden of one's own cultural experience" (Robinson). Ashcroft delves into the challenges that the colonized countries face, further advancing the exploration of the relationship between the colony and the colonized.

All in all, the postcolonial theory dedicates its effort to the discussion of the relation between the colonized and the colony through the time span during the imperialized history and the eras that follow. With Edward Said’s work about inequality, narrative, and misperception in the colonial experience as the basis, Bhabha added on a clearer investigation of the empire and how its former colonies perceive each other, as well as the reason behind that perception. Spivak added a more detailed perspective from the ones being colonized, illustrating the barriers that they face. Ashcroft was dedicated to the abrogation and appropriation within the colonial body, further completing the understanding of an imperialistic experience.

The Japanese Colonization Historiography

Japan’s imperialization started in 1845, alongside the meiji restoration, which greatly strengthened the country. Japan’s first colony was the southern islands of Sakhalin and the Kuril Archipelago, soon expanding to other countries in Asia. Most of the colonized countries share the characteristics of having a close geographical location, often also meaning a similar cultural context. Two of Japan’s most influential acts of colonization would be the establishment of colonies in China and South Korea.

During the imperialistic periods, the Japanese government implemented measures to spread the nationality and support for the imperialization through the strategic design of related textbooks, which made instrumental contributions to the establishment of national civil support for the imperialization (Yoshiko and Hiromitsu). Japanese government emphasized the slogan that their rule is the “Governance of the Kingly Way”, and that all citizens should “follow the way of Heaven and the people will be in peace” (Agnew). As observed by Ashcroft, these narratives were strategically planned in a manner that favored the colonizing empire.

As the empire expanded, the Japanese government developed the justification for its action with the name of creating a “Eastern Asia Co-Prosperity Sphere”, which supposedly brought benefits to its colonies. Using China as an example, figure 1 below shows that Japan did succeed in becoming the country which China imports the most from, taking a percentage of 24.51% in the period in which most of Japanese colonies in China lasted. However, further research also shows that there hasn’t been any actual improvement in the economic output of China in the period which it was ruled by Japan (Grajdanzev).

Country	%
Japan	24.51
United States	22.05
Great Britain	17.02
British India	9.65
Germany	4.19
Java (Dutch East Indies)	2.91
French Indochina	2.39
Russia (Soviet Union)	2.10
Belgium-Luxembourg	1.85
Singapore	1.56
Australia	1.25
Other countries	10.53

(Keller and Shiue)

Fig 1. Sources of Chinese Imports from 1900 to 1946

The rule of China started with the breakthrough of the Mukden incident in 1913¹, taking control of Manchuria. Following Japanese occupation, a series of immigration followed, mostly for the Manchuria labor services, and primarily in the field of mine and road construction (Grajdanzev).

With the colonizing process, the benefit of increased job opportunities and advanced technologies such as railway constructions were brought to China. In the meantime, the imperialistic control coexisted with numerous violent military conflicts as Japan aimed to expand its control within China. After the conquer of Nanjing, which used to be the Nationalists Chinese's capital, the mass killing from the Japanese Imperial Army resulted in the deaths of 100,000 to more than 300,000 civilians. Historically known as the Nanjing Massacre, the genocide commanded by the Japanese commanding general Matsui Iwane, left the country with inescapable trauma of the imperialistic history (Lu).

The Japanese colonization of South Korea then started in 1910, where Japan came in to replace the monarch government with a puppet government, along with a series of other changes that imposed an altered version of their identity. Action for this purpose included banning the Korean languages in schools and strategically designing Korean textbooks. This is a clear demonstration of how the Japanese empire, just like the western colonies, establishes their rule with strategically altered narratives for the colonized country. On top of the colonial governance, the use of the Korean workforce also left inescapable influences, where the Korean workers were exploited with disregards of human rights, racially discriminated, and paid extremely low salary (Robinson).

Historical comments that highlight the benefits that Japan brought during its rule are definitely present, such as how David Landes argued that the positive performance of South Korea after the colonial rule is proof of the Japanese empire's accomplishment. In the meantime, there are also opinions that believe the period of struggle for South Korea's economics during the 1930s to 1950s is another proof that the Japanese empire's support of economics was insufficient (Booth and Deng).

For both China and South Korea, the Unit 731 executed horrific experiments on civilians from both countries, and the war crimes of comfort stations also applied to both. Unit 731 was a troop officially named Epidemic Prevention and Water Purification Department of the Kwantung Army, established in the mid 1930s as an institution aimed to develop biological weapons for the Japanese army. Within the decades of Japan's occupation, Japan conducted experiments of lethal biological weapons experiments on at least hundreds, at most half-million of Chinese civilians from northeastern China and Korean civilians from the northern side of Korea peninsula and colonized Asians from several other locations. The experiment site was destroyed after Japan was bombed twice in 1945 and began its surrender, and hundreds of the prisoners were killed.

Additionally, the comfort stations were established along the front lines of the Japanese Imperial Army. As Harvard Review records: "Euphemistically labeled 'comfort women,' between

¹ On September 18th of 1931, a portion of Manchuria's Railway track was blown off. Despite the Chinese government's appeal to the League of Nation, Japan started its intruding occupation of the Manchuria region by establishing its puppet government (Ferrell).

80,000 and 200,000 women are estimated to have been coerced into supplying sex for Imperial Japanese soldiers during WWII” (Chang). The women primarily came from Korea, but also from China, and a range of other imperialized countries. Many of the women were made to believe that they were recruited for a factory job or a nursing job, when in fact they were forced into providing sexual service for the Japanese soldiers. Most of the women in the comfort stations died without escaping, while most of the ones who managed to survive also couldn’t publicly share their experience because of social stigma (Chang).

The San Francisco Peace Treaty in 1952 was politically viewed as a resolution of the war crimes, but it was not viewed as a satisfactory response for many civilians and many colonized countries, namely including South Korea and China (Miller).

After the official surrender from Japan, the Tokyo trial was held in 1946 to settle the charges against Japan for its actions during the WWII. Japan was judged to be guilty, with a number of Japanese officials being charged to death or to a life in prison (The National WWII Museum). However, the trial is still often criticized for the absence for a number of political leaders, dispute between the judges, and how the judgement was “victor’s justice” (Saito).

As decades have passed, the influence of the colonial memory has not diminished. Further evaluations on the current East Asian political standings and societal perceptions provides the understanding of how the colonial past sheds its shadow in the modern world.

The Postcolonial Results in Modern Times

Since the end of Japan’s imperialization along with its surrender in 1945 to end the Second World War, the political dynamics in Asia and the rest of the world has endured a series of change. Nonetheless, the historical memories from the colonization remain in the dialogues of civilians and governments’ official speakers, primarily being a topic of discussion in China and South Korea, requiring feedbacks from Japan, and occasionally including the United States government². The following section will examine the common opinions from China, Japan, and South Korea as a comparison with the traditional post-colonial theories.

From the perspective of South Korea, on the societal level, “the trauma of the colonial time is still alive” (Korostelina). In various interviews with Korean citizens from different social backgrounds, keywords regarding the “cruel and unjust” actions of the Japanese rule, and “deep grievances [concerning the way in which] its history was distorted and deliberately wiped by Japan” was most frequently mentioned (Korostelina). Although there are opinions that impose the Korean government to share part of the responsibility, the mainstream idea still focuses on the hatred towards Japan. On the political layer, the historical disputes remain unresolved, mostly on the possession of the Dokto Island and the war crimes.

In the meantime, there is still frequent friendly communications in-between. The political trend of befriending with Japan has come to be progressing more so since the US started its Indo-Pacific Strategy, one which will be greatly benefited if South Korea and Japan can work

² Joe Biden hosted a three-way conference with leaders from Japan and South Korea. The primary aim is to establish US’ allies in Asia, which also included addressing the dispute for the colonial past (Harold).

jointly. From the economical metric, the cooperation between Japan and South Korea is also influential. Being one of the major countries of foreign direct investments with South Korea, Japan comes in third from all the countries' amount of OFDI (Outward foreign direct investment) with South Korea (Kim and Mah). Indeed, it can be concluded that South Korea's attitude is "ambivalent", a "mix of attraction and repulsion", just as Homi Bhabha commented to be. However, it still shows a difference with the attraction of admiration for Japan's more modernized statues, as presented in the postcolonial theories. Instead, most of the cooperation, or in other words, the "attraction", came from political strategies based on the current political dynamics. In the memory of the imperialistic history, the "repulsion" of hatred in Japan's cruelty takes the lead.

For China, the common societal view is still working against Japan, becoming especially sensitive when coming to the historical past. It is especially sensitive in Chinese social media when encountering topics of Japan's imperialistic history; the Japanese Prime Minister Abe's visit to the Yasukuni Shrine were strongly condemned (Shibuichi). Mostly, the government's political layer also showed firm stance of negating the Japanese's any act of denying its responsibility for the WWII war crimes. In 1972, a speech from a national noted that the government "firmly opposes the revival and outward expansion of Japanese militarism" ("Full Text of Chinese President's Speech at Commemoration of 70th Anniversary of War Victory"). In the face of Japan's act of attempting to strategically plan the narrative in its textbooks, it was reported that both the Chinese government and people were "strongly indignant about and dissatisfied with the new Japanese history textbook for the year 2002 compiled by right-wing Japanese scholars" (Masalski).

In reference to the expected "ambivalence" between the colony and the colonized, there have been scholarly works that referred to China's attitude towards Japan to be a mix of "admiration, envy, and occasional hostility", often also described with "contradiction of deep distrust on one hand and important interest on the other" (Smith). Under such complexity, it concludes that Homi Bhabha's analysis about the colony and colonial subject's relation does apply to China and Japan.

From Japan, the strategic narrative of the imperialistic history extends to today, although creating a smaller impact. Early as the 1990s, Fujioka Nobukatsu, a professor of education at Tokyo University, started his advocates to correcting Japan's history textbooks to avoid the "dark history". Later in 2002, as the government explains, they believe the comfort women's dispute is not proven by any official documents, and hence it is unreasonable to include this event in the textbook. With such wordings, the Japanese government removed the contents about comfort women from their official textbook. "It has been a matter of recovering national pride", was also a common reason that Japan government used to explain this change (Chizuko and Sand). A Japanese historian, Richard H. Minear, once illustrated: "As a practicing historian, I encounter at every turn the power textbooks exercise over my students' minds" (Masalski). Clearly indicating how Japanese government's act will be leading to great difference in the country's civilians

recognition of its imperial past, possibility creating great setbacks to the progress of revealing the historical facts.

Namely, it should be noted how such acts provoked dissatisfaction even amongst the Japanese society. A group of international scholars related to Japan or other former colonies jointly expressed their protest against the 2002 version of textbook, elaborating that “it negates both the truth about Japan's record in colonialism and war and the values that will contribute to a just and peaceful Pacific and World community” (Masalski). There were also a range of other objections from a range of other organizations and individuals. In the end, the protests were effective: only 0.03% of the Japanese junior high schools adopted the 2002 version of historical textbook (Loewen).

Noting the failure of the attempt to display the strategically planned representation within the country, it reflects how Japan, as a former empire, did present its efforts in altering the imagery of the colonized and its own past. Another difference between Japan's misrepresentation and the usual narratives under postcolonialism is how there weren't any reported difficulties in understanding Chinese or Korean culture. In fact, there have been studies of Chinese ancient and modern literature since 1920, further lowering the barrier of understanding between these foreign cultures (Maruyama).

Specifically in terms of the issue of comfort women, Spivak's analysis of the silenced expression does show to apply suitably. While the comfort station remains one of the most condemned war crimes, the process for the establishing the social recognition of the event took an incrementally long route. One of the major challenges to the process of revealing the integral view for this past is the social stigma on women for losing their “loyalty” to their husbands, treating the experience of being a “comfort women” as an embarrassing past which they do not want to bring up again. Just as how Spivak presented her opinions, the comfort women, being a community that was harshly treated under the imperial rule, did encounter challenges because of the gender role and the position that the colonial history provided.

The first victim that spoke publicly to share her experience was a Korean women named Kim Hak-Sun, who spoke to the reporters about her story as a “comfort woman”. Kim expressed herself to have made the decision to take about this past that was insufferable to just think about because she couldn't stand the government's lies (Yōko). Following Kim's break of silence, many other comfort women have stood out as alibi of the accusation for Japanese Imperial Army's crime or by providing additional evidence. Afterwards, the global recognition of the comfort women community endured a further increase, and civil organizations such as the Kawasaki Omoni no Kai (Korean Mother's Association in Kawasaki) and the Meari Kai (Association Echo) all promoted the public recognition of the comfort women (Seo).

In response to such connotations, the Japanese government also demonstrated varying attitude, depending on the specific official. In 1993, the Chief Cabinet Secretary Kono Yohei once stated: “The Government of Japan would like to take this opportunity once again to extend its sincere apologies and remorse to all those ...who suffered immeasurable pain and incurable physical and psychological wounds” (Easley). However, the later version of the Japanese

Diplomacy notes against apologizing for its war crimes, and the Prime Minister Abe's visit to the Yasukuni Shrine once again diminishes the international society's hope towards Japan's apology (Korstelina). While the Chinese and South Korean government and civilians all showed discontent regarding such acts from the Japanese government, the Japanese government still proceeded with their own acts that led further away from apologizing. Acts of such provide proof of Spivak's work about how the colonizing country made its effort in avoiding the parts of history which doesn't favor their narrative.

On the political layer, the building of occasional cooperation of Japan, South Korea and China should still be addressed. Japan's prime minister Abe has been noted to aim for building "strategic bilateral ties for mutual benefit" with China; in October 2006, Abe was also the first post-war prime minister that chose visit China as his first foreign visit (Smith). Fukuda's later visits and works is commented to be expanding on Abe's will (Smith). With the Korea-Japan relation, there has also been a strong economical tie, with over 9 million visitors per year and billions of output each year (Easley). Though with Moon Jae-in and Shinzo's contacts, the two countries relation has been commented to reach its worst since the normalization of their communication in 1965 (Sakaki).

Comparison With Traditional Postcolonial Theory

As a non-western colony, the Japanese empire also worked through its expansion with the name of brings prosperity to all, just as the post imperialization theory expected. On the other hand, the strategy of creating inequality in Japan's plan was different, stemming out of the geographical and cultural similarities. With the approach of creating an "Eastern Asia Co-Prosperity Sphere", Japan's justification showed slight differentiation from the separation of "self" and "other" in the western empires and focuses on the unity and harmony with imagined blood and organic ties between Japan and its colonies (Park). Though there is still the identification of the inferiority for the colonies; specifically showing in events such as exploited labor that specifically discriminates on workers from its colonies.

Another major difference with the postcolonial ideas is in the contrast with Ashcroft's ideas about abrogation and appropriation, demonstrated in how there wasn't much evidence about Japan's strategic misrepresentation of South Korea or China's culture. This could have been because of the similar cultural context, which Japan used to encourage support from the colonized countries. Hundreds of years before the Japanese imperialization, the frequent communication in the Chinese Tang dynasty and various other cultural communication opportunities provided a higher level of similarity between the three countries.

Further extending on the cultural exchanges in the modern times, Japanese society's demonstrated interest towards Chinese literature also limits Japan's abrogation of China's culture. This also meant that the scenario of the colonizer being unable to understand the colony's local culture also isn't present.

Nonetheless, there were still acts where Japan attempted to replace the Korean and Chinese language with Japanese, the twists of historical narratives also served as proof that the

intentional narrative still takes a major role in Japan's imperial past. After Japan's rule, it also left Korea with an influence in its food culture, additional words in the language, and other perspectives of cultural changes. Such evidence contributes to explaining how both narrative and abrogation of native culture still played a role in Japan's imperialization.

Although it's still worth noting that the transformation of language in Korea was not successful, and as Koreans regained their usage of their native language, the barrier of expression diminished. Therefore, the "subaltern", which would be the Koreans, did not face the expected challenge of only being able to express themselves under the foreign framework. Building upon this, an alternate definition of the "subaltern" as woman, who were less respected in the society brings another conclusion to address the challenges that women uniquely face in the postcolonial world. Specifically, regarding the victims of the comfort station, being women, who faced the additional boundary from expression, coming from the social stigma implied for the past of vast sexual harassment and rape. Directly relating to how the social structure and gender expectation in the postcolonial world similarly imposes its impact on the victims from the imperialistic times.

During Japan's imperial rule, another key would be the altered narrative with its acts of rewriting the textbooks in both their own country and their colonies for justification and propaganda of their act. As Edward Said explains, imperialism necessarily brings the stage that "we give you our history, we will write it for you, we will reorder the past." This also comes to be similar to the opinion from Ashcroft, the narratives were strategically planned in a manner that favors the colonizing empire. The rewriting of their colonies' history serves as a representative example of how the Japanese empire follows both Homi Bhabha and Bill Ashcroft's expectations.

Commenting on the relation between Japan and its former colonies, the ambivalence is definitely playing a key role. As previously described, China's attitude with Japan since the imperialistic experience has been a complex result of "admiration, envy, and occasional hostility". Similarly for Korea, it also cannot be simply defined as "attraction" nor "repulsion". The main difference for the ambivalence in the Japanese empire is shown in how the worship of the colonizing country did not seem as obvious in Japan's colonization, instead the hostility of the former colonizer seemed to be more mainstream. Nonetheless, disregarding the minor variations, the theory is still applicable. To encapsulate, non-western powers also encounter a similar experience as western empires, which stems from the shared characteristic to endure a history which the empire is narrated to be superior to the colonized.

Policy Recommendations

As there are still a copious number of topics regarding colonial history that comes unresolved amongst the involved countries, efforts devoted to the compliance of such topics comes with priority to be resolved. In comparison with the common dispute among civilians regarding such topics and conflicting stances between government speakers, a guided discussion between countries' leaders fosters the route to a more organized and peaceful international

relation. The agreement between governments that gives all perspectives a satisfactory result is the first step for the dispute of history to be settled.

Within the dialogues, the main issue to be addressed is comfort women, forced labor and unit 731 for both China and South Korea; Nanjing massacre in China should be the focus. Acknowledging that there have been attempts to reach the accord from various perspectives, including but not limited to the San Francisco Peace Treaty, conferences led by the United States to resolve the conflict between South Korea and Japan, the 2015 Comfort Women Agreement and several others (Se Eun Gong). The agreement reached in 2006 when Abe visited China could also be used as a reference in this case (Zhang). Nonetheless, most of society were not content with the resolution from these events, the demand for an official and sincere apology from the Japanese government comes to be the core of what remains disputed.

With an integrated consideration of different opinions, the apology from Japan will be the optimal conclusion of the disputed affairs. As many scholars have inferred, under the framework of human rights in the current society, “individual rights were not addressed” in Japan’s original response in 1965. From the Japanese government, they continuously “apply the same international law and history but are continuing to clash over the issue of ownership” (Korostelina). Apology from Japan for the war crime dispute not only provides the victims and the former colonies an official response, but it also serves to end the use of such events as a political weapon against Japan in the future, especially as it is commonly agreed on the criticism of Japan’s war crimes committed during the Second World War. From the Japanese government’s consideration, the United States, which Japan have been actively engaging to build a positive relation with, has also been pressuring Japan along with Canada and a series of European countries in working towards an official apology for its war crimes.

It is already clear that Japan’s past approach of resolving the problem with money has failed, as quoted from a South Korean citizen, also a former victim of Japan’s war crimes: “I would rather starve to death than take that money” (Chang). With Prime Minister Abe’s statements of not apologizing to the comfort women because the comfort station cannot be historically proven with governmental documents, even many of the Japanese citizens recognizes as “a source of embarrassment”. Noting that there have been government officials before that have agreed to the apology, and that there is a trend of public opinion, showing in how “three Japanese city councils themselves have passed motions calling for a more direct apology”, there is a feasibility in formatting an apology from the Japanese side (Yōko). All in all, the sincere apology from Japan would bring the maximized benefit for all.

Further on, the textbook is also a key tool that needs to be aligned between countries in order to reach a stable consensus among the civilians. As historians Laura Hein and Mark Selden note: "history and civics textbooks in most societies present an 'official' story highlighting narratives that shape contemporary patriotism" (Masalski). Setting an official agreed perception of Japan’s committed past in its occupation should egress along with the settled agreement governments.

Hence, it comes to the conclusion that the extension of open dialogues revolving around imperialistic history is one of the most important approaches to the traumatic experience, with an extension to aligning historical textbooks to reach a through consent in the society.

Theoretical Recommendations

As proven, there are still missing pieces in the scholarly investigation of the interaction between Japan and its former colonies. With the unique Asian cultural identities and the neighboring geographical location, parts of the traditional postcolonial theory fail to show its application with Japan's imperialistic history. Hence, it is reasonable to infer that there should still be further investigation in the field of postcolonial interaction patterns for non-western empires, as the current theories are not suitable for the empires that targets nearby culture and location.

Specifically in Japan, it also seems that the recognition of the imperialistic history is rather limited. There seems to be investigation of the Meiji reform in the lens of imperialization; it is crucial to realize how the Meiji period cannot be isolated as an island history and should be considered with its international influence (Schmid). Although understanding that it is a common scenario for countries to pay less attention to the shameful or the criticized parts of their history, it is still necessary for the country of Japan to bring forth more support with such investigations using local resources, which is a crucial perspective to reveal a holistic view of the imperialistic experience.

As much as western centrism is a commonly discussed and investigated topic, there is still space for more research to develop more comprehensive understanding on the colonization within Asia, or possibility other specific region.

Conclusion

As above has shown, the Japanese Empire, being a representative example of a non-western empire, is one which the classical postcolonialism theory applies only to an extent, which the similarities namely exclude the emphasis on cultural abrogation and appropriation. As explained, further academic research will be greatly beneficial to fill in the missing components of the scholarly studies of non-western empires, possibly coming to a more integrated conclusion on the result that the difference in the empires' culture may lead to.

Since the second world war, the world seems to have stepped away from imperialization, some would even regard it as a past not worth of reflection. I believe it is worth reminding that the menace of revisiting imperialism has not been terminated; looking forward, with the spread of US and China's influence, stronger countries' control of weaker states could happen again. In such a possible future, the fully examined experience of not only western, but also non-western empires could be instrumental to the understanding of a new world standing in the future. Eventually, the imperialistic past most definitely left its visible traces to our current world, and the unhealed scars left spaces for conflicts that are continuing even today.

Works Cited

- Agnew, Junko. "The Politics of Language in Manchukuo: Hinata Nobuo and Gu Ding." *Modern Asian Studies*, vol. 49, no. 1, 17 June 2014, pp. 83–110, <https://doi.org/10.1017/s0026749x13000541>.
- Ashcroft, Bill, et al. *The Empire Writes Back : Theory and Practice in Post-Colonial Literatures*. Brantford, Ont., W. Ross Macdonald School Resource Services, 2011.
- Austin, Allan. "Superman Goes to War: Teaching Japanese American Exile and Incarceration with Film." *Journal of American Ethnic History*, vol. 30, no. 4, 2011, p. 51, <https://doi.org/10.5406/jamerethnhist.30.4.0051>. Accessed 8 Nov. 2020.
- Berger, Mark. "Decolonisation, Modernisation and Nation-Building: Political Development Theory and the Appeal of Communism in Southeast Asia, 1945-1975." *Journal of Southeast Asian Studies*, vol. 34, no. 3, 2003, pp. 421–448, www.jstor.org/stable/20072531.
- Bhabha, Homi. "Of Mimicry and Man: The Ambivalence of Colonial Discourse." *October*, vol. 28, no. 1, 1984, pp. 125–133, <https://doi.org/10.2307/778467>.
- Boehmer, Elleke. "Transfiguring: Colonial Body into Postcolonial Narrative." *NOVEL: A Forum on Fiction*, vol. 26, no. 3, 1993, p. 268, <https://doi.org/10.2307/1345836>.
- Booth, Anne, and Kent Deng. "Japanese Colonialism in Comparative Perspective." *Journal of World History*, vol. 28, no. 1, 2017, pp. 61–98, www.jstor.org/stable/44631508. Accessed 3 Sept. 2023.
- Chang, Mina. "The Politics of an Apology: Japan and Resolving the "Comfort Women" Issue." *Harvard International Review*, vol. 31, no. 3, 2009, pp. 34–37, www.jstor.org/stable/42763319.
- Chizuko, Ueno, and Jordan Sand. "The Politics of Memory: Nation, Individual and Self." *History and Memory*, vol. 11, no. 2, 1999, p. 129, <https://doi.org/10.2979/his.1999.11.2.129>.
- Collier, James. "The Theory of Colonization." *American Journal of Sociology*, vol. 11, no. 2, 1905, pp. 252–265, www.jstor.org/stable/2762664.
- Derby, Orlo. "Japanese and American Stereotypes." *The Clearing House*, vol. 33, no. 1, 1958, pp. 7–8, www.jstor.org/stable/30190624. Accessed 23 Aug. 2023.
- Easley, Leif-Eric. "Korean NGOs and Reconciliation with Japan." *Journal of East Asian Studies*, vol. 23, no. 1, 10 Feb. 2023, pp. 45–70, <https://doi.org/10.1017/jea.2022.21>. Accessed 4 May 2023.
- Ferrell, Robert H. "The Mukden Incident: September 18-19, 1931." *The Journal of Modern History*, vol. 27, no. 1, 1955, pp. 66–72, www.jstor.org/stable/1877701.
- "Full Text of Chinese President's Speech at Commemoration of 70th Anniversary of War Victory." www.fmprc.gov.cn, www.fmprc.gov.cn/eng/topics_665678/2015zt/jnkzsl70zn/201509/t20150903_705223.html. Accessed 26 Aug. 2023.
- Goldar, Bishwanath, and Etsuro Ishigami. "Foreign Direct Investment in Asia." *Economic and Political Weekly*, vol. 34, no. 22, 1999, pp. M50–M60, www.jstor.org/stable/4408017.

- Grajdanzev, A. J. "Manchuria as a Region of Colonization." *Pacific Affairs*, vol. 19, no. 1, Mar. 1946, p. 5, <https://doi.org/10.2307/2752218>. Accessed 19 Nov. 2019.
- Harold, Scott W. "The Japan-U.S. Summit and Cooperation with South Korea." *Rand.org*, 2021, www.rand.org/blog/2021/04/the-japan-us-summit-and-cooperation-with-south-korea.html. Accessed 4 Sept. 2023.
- Johnson, Chalmers. "How China and Japan See Each Other." *Foreign Affairs*, vol. 50, no. 4, 1972, p. 711, <https://doi.org/10.2307/20037943>. Accessed 10 Dec. 2019.
- Keller, Wolfgang, and Carol H. Shiue. "China's Foreign Trade and Investment, 1800-1950." National Bureau of Economic Research, 1 July 2020, www.nber.org/papers/w27558. Accessed 23 Aug. 2023.
- Kim, Eun Mee, and Jai S. Mah. "Patterns of South Korea's Foreign Direct Investment Flows into China." *Asian Survey*, vol. 46, no. 6, Dec. 2006, pp. 881–897, <https://doi.org/10.1525/as.2006.46.6.881>. Accessed 26 Apr. 2020.
- Korostelina, Karina. "The Normative Function of National Historical Narratives: South Korean Perceptions of Relations with Japan." *National Identities*, vol. 21, no. 2, 29 Nov. 2017, pp. 171–189, <https://doi.org/10.1080/14608944.2017.1401599>. Accessed 21 Aug. 2020.
- Krueger, Justin. "To Challenge the Settler Colonial Narrative of Native Americans in Social Studies Curriculum." *The History Teacher*, vol. 52, no. 2, Feb. 2019, pp. 291–318.
- Loewen, James. *Lies My Teacher Told Me: Everything Your American History Textbook Got Wrong*. Atria Books, 16 Oct. 2007.
- Lu, Suping. "The Nanjing Massacre: Primary Source Records and Secondary Interpretations—a Textual Critique of Bob Tadashi Wakabayashi's Review." *China Review International*, vol. 20, no. 3/4, 2013, pp. 259–282, www.jstor.org/stable/43818315. Accessed 18 Mar. 2022.
- Maggio, J. "'Can the Subaltern Be Heard?': Political Theory, Translation, Representation, and Gayatri Chakravorty Spivak." *Alternatives: Global, Local, Political*, vol. 32, no. 4, 2007, pp. 419–443, www.jstor.org/stable/40645229. Accessed 24 Aug. 2023.
- Marotta, Vince. "Cultural Hybridity." *The Blackwell Encyclopedia of Sociology*, 22 Oct. 2020, pp. 1–4, <https://doi.org/10.1002/9781405165518.wbeos1616>.
- Maruyama, Noboru. "Studies of Modern Chinese Literature in Japan." *Modern Chinese Literature Newsletter*, vol. 2, no. 2, 1976, pp. 1–15, www.jstor.org/stable/42769596. Accessed 23 Aug. 2023.
- Masalski, Kathleen. "Examining the Japanese History Textbook Controversies." *Indiana University Bloomington*, Nov. 2001, www.indiana.edu/~japan/Digests/textbook.html.
- Miller, Jennifer M. *Cold War Democracy. The United States and Japan*. 2023. Cambridge, Harvard University Press, 2019.
- Nasrullah Mambrol. "Ambivalence in Post-Colonialism." *Literary Theory and Criticism*, 27 Sept. 2017, literariness.org/2017/09/27/ambivalence-in-post-colonialism/.
- Park, Hyun Ok. "Korean Manchuria: The Racial Politics of Territorial Osmosis." *South Atlantic Quarterly*, vol. 99, no. 1, 1 Jan. 2000, pp. 193–217, <https://doi.org/10.1215/00382876-99-1-193>. Accessed 31 Aug. 2020.

- Robinson, Heather. *Creole Composition: Academic Writing and Rhetoric in the Anglophone Caribbean*. Edited by Vivette Milson-Whyte et al., Parlor Press, 2019.
- Saito, Hiro. “The Legacy of the Tokyo Trial.” JSTOR, University of Hawai’i Press, 2017, www.jstor.org/stable/j.ctt1wn0r56.9. Accessed 8 Mar. 2023.
- Sakaki, Alexandra. “Japan-South Korea Relations—a Downward Spiral: More than “Just” Historical Issues.” SWP (Stiftung Wissenschaft Und Politik) , 2019, www.swp-berlin.org/fileadmin/contents-/products/comments/2019C35_skk.pdf.
- Schmid, Andre. “Colonialism and the “Korea Problem” in the Historiography of Modern Japan: A Review Article.” *The Journal of Asian Studies*, vol. 59, no. 4, Nov. 2000, pp. 951–976, <https://doi.org/10.2307/2659218>. Accessed 3 June 2019.
- Schmidt, Volker. “How Unique Is East Asian Modernity?” *Asian Journal of Social Science*, vol. 39, no. 3, 2011, pp. 304–331, www.jstor.org/stable/43498790. Accessed 23 Aug. 2023.
- Se Eun Gong. “U.S. Allies South Korea and Japan Move Closer to Resolve Forced Labor Feud.” NPR, 6 Mar. 2023, www.npr.org/2023/03/06/1161250054/u-s-allies-south-korea-and-japan-move-closer-to-resolve-forced-labor-feud.
- Seo, Akwi. “Toward Postcolonial Feminist Subjectivity:: Korean Women’s Redress Movement for “Comfort Women.”” JSTOR, University of Hawai’i Press, 2018, www.jstor.org/stable/j.ctv3zp07j.21.
- Shibuichi, Daiki. “The Yasukuni Shrine Dispute and the Politics of Identity in Japan: Why All the Fuss?” *Asian Survey*, vol. 45, no. 2, 2005, pp. 197–215, www.jstor.org/stable/10.1525/as.2005.45.2.197, <https://doi.org/10.1525/as.2005.45.2.197>. Accessed 8 Dec. 2019.
- Shilliam, Robbie. “Modernity and Modernization.” *Oxford Research Encyclopedia of International Studies*, 1 Mar. 2010, <https://doi.org/10.1093/acrefore/9780190846626.013.56>.
- Smith, Paul J. “China-Japan Relations and the Future Geopolitics of East Asia.” *Asian Affairs*, vol. 35, no. 4, 2009, pp. 230–256, www.jstor.org/stable/30172699.
- The National WWII Museum. “Tokyo War Crimes Trial.” The National WWII Museum | New Orleans, www.nationalww2museum.org/war/topics/tokyo-war-crimes-trial. Accessed 4 Sept. 2023.
- Yōko, Hayashi. “Issues Surrounding the Wartime “Comfort Women.”” *Review of Japanese Culture and Society*, vol. 11/12, 1999, pp. 54–65, www.jstor.org/stable/42800182.
- Yoshiko, Nozaki, and Inokuchi Hiromitsu. “Japanese Education, Nationalism, and Ienaga Saburo’s Court Challenges.” *Bulletin of Concerned Asian Scholars*, vol. 30, no. 2, June 1998, pp. 37–46, <https://doi.org/10.1080/14672715.1998.10411042>.

A 3-tiered Color Scale Nitrate Biosensor By Dongmin Choi

Dongmin Choi¹, Yejun Hwang², Grace Lee¹, Sunghyun Yoon³, June Kim³, Faith Dallao⁴, Jeongmin Lee⁵, Juhan Kim⁶, Minwoo Chung⁷, June Hyuk Kwak⁸, Jaden Cho⁹, Chaeyoung Sohn¹⁰, Eileen Yoo¹¹, Sungho Yoon⁶, Hyunji Yang¹², Jackie Lee¹³, Ryan Lee¹⁴

¹ Chadwick International School, 45, Art center-daero 97beon-gil, Yeonsu-gu, Incheon, 22002, Republic of Korea

² Fayston Preparatory School, 11, Dongcheon-ro 358beon-gil, Suji-gu, Yongin-si, Gyeonggi-do, Republic of Korea

³ North London Collegiate School Jeju, 33, Global Edu-ro 145beon-gil, Daejeong-eup, Seogwipo-si, Jeju-do, 63644, South Korea

⁴ Seoul Foreign School, 39 Yeonhui-ro 22-gil, Seodaemun-gu, Seoul

⁵ Seoul Scholars International Art & Design, 224 Seocho-daero, Seocho-gu, Seoul, 06655, Republic of Korea

⁶ St. Mark's School, 25 Marlboro Road Southborough, MA, 01772, United States

⁷ The Masters School, 49 Clinton Ave, Dobbs Ferry, New York, 10522, United States

⁸ Saint Paul Preparatory Seoul, 14-8, Seochojungang-ro 31gil, Seocho-gu, Seoul, 06593, South Korea

⁹ Asheville School, 360 Asheville School Road, Asheville, North Carolina, 28806

¹⁰ Laurel Springs School, 302 El Paseo Rd, Ojai, CA 93023, United States

¹¹ 'Iolani school, 563 Kamoku St, Honolulu, HI 96826, USA

¹² Westminster School, 995 Hopmeadow St, Simsbury, CT 06070, United States

¹³ Kent School, 1 Macedonia Rd, Kent, CT 06757, USA

¹⁴ Western Academy Of Beijing, 10 Laiguangying East Rd. Chaoyang District, Beijing 100102, China

Abstract

Water is necessary to sustain life. Yet, the unprecedented effects of industrialization have increased the concentration of inorganic minerals like nitrogen in rivers and lakes, accelerating eutrophication around the world. This hazard comes without notice, silently transforming pristine water into lifeless, deoxygenated dead zones. Subsequently, the dead bodies of creatures and toxins secreted by algae contaminate the water, requiring layers of purification for humans to use. To address this issue, multiple attempts have been made to monitor the inorganic minerals to prevent eutrophication. The monitoring requires a sensitive sensor that can detect nitrate levels in water, and one type of sensor that has gained popularity is a biosensor. Conventional nitrate biosensors employ a relatively simple approach in using a single type of reporter. In contrast, we selected two types of reporters packed in an alginate bead to effectively visualize the nitrate level through a color change. This color scale not only offers a more intuitive way to gauge ion concentration but also allows for a wider range of measurements compared to when using a

single color. Moreover, we successfully controlled the amount of NsrR in *E. coli*, a key factor in nitrate sensing, by testing with various expression units, verifying the sensor's potential to detect nitrate in a wide range of concentrations. We established a novel approach for whole-cell biosensors that can express the nitrate level through a dynamic color change in a sustainable and cost-efficient manner.

Introduction

Eutrophication occurs when excess nutrients accelerate algal growth, leading to the reduction of dissolved oxygen and the death of aquatic species (Chislock, 2013). The root causes of eutrophication involve human activities, such as agriculture, industrial production, and sewage disposal, as they excessively increase the concentration of nitrate in bodies of water (NOAA, 2017).

While the growing trend of more devastating cases of algae growth has been undeniable, the issue of eutrophication entered the conversation of public safety only recently. Specifically, the issue of algae entering domestic and agricultural water supplies has sparked serious concerns in South Korea. Nakdong River, the longest river and water source for 13 million, has seen record levels of toxic blue-green algae growth in the last 5 years. At a farm in Yangsan, a city near the river, the irrigated water was found to contain 5079 $\mu\text{g/L}$ of microcystin, which is at least 20 times more toxic than cyanide and 635 times higher than the U.S. Environmental Protection Agency's recreational water advisory limit of 8 $\mu\text{g/L}$ (Nakdong River Network, 2022). The economic impact is also significant, as local products produced near the Nakdong River may be instilled with a negative perception by the public, impacting the local industry.

Although there are numerous methods for restoring damaged ecosystems through chemicals like copper sulfate (CuSO_4), as well as physical methods like sediment dredging, the efficacy and accessibility of such methods are limited, especially in rural areas. Algal blooms are influenced by several abiotic factors, such as temperature, sunlight, and nutrient availability, but controlling temperature and sunlight is challenging. Thus, more effective solutions focus on preventive measures than reactive treatments, monitoring and controlling nitrate levels. For accurate insights, it is crucial to assess the entry points and quantities of nitrate, but the scarcity of affordable nitrate sensors poses a significant challenge, hindering widespread installation.

Indeed, the market availability of these biosensors is still relatively limited compared to other types of nitrate sensors. Nonetheless, current nitrate biosensors on the market have several advantages. First, nitrate biosensors have a high degree of specificity towards nitrate ions, which reduces the interference from other compounds. This guarantees precise and trustworthy measurements, even in water samples rich in confounding ions, such as phosphates. Additionally, a lot of nitrate biosensors are made to be easily manipulated and portable, which makes it easier to conduct tests on-site and minimizes the need for sample transportation and laboratory analysis. Finally, nitrate biosensors may occasionally be less expensive than conventional laboratory-based techniques, especially when used for high-frequency testing or long-term

monitoring (Almeida et al., 2010). Therefore, portable and cost-effective nitrate biosensors have strong competitive potential in the market.

Previous research created a nitrate biosensor using *E. coli* and *Enterobacter cloacae* to express ice nucleation (*inaZ*) and green fluorescent protein (GFP) reporter genes under the control of a nitrate-responsive promoter (DeAngelis, Kristen M et al.). This allowed the bacteria to serve as sensors, detecting and indicating nitrate levels in their environment by producing measurable ice nucleation activity and GFP fluorescence. Similarly, there are developments in nitrite biosensors by employing nitrite reductase enzymes (Almeida, M Gabriela et al.), such as cytochrome c nitrate reductase (*ccNiR*) from *Desulfovibrio desulfuricans*, integrated into electrochemical and optical transducers to translate the enzymatic reduction of nitrite into measurable signals. Both research leverage specific gene expression and enzyme activity to provide precise and rapid detection of nitrogen compounds. While significant progress has been made in detecting nitrogen compounds, there hasn't been much effort to use a color scale to intuitively express the nitrate level in bodies of water.

Hence, we delved into a novel approach to expressing nitrate levels in a 3-tiered color sensor. The proposed sensor consists of two reporter genes, *vsfGFP* and *mScarlet*, with the Lon protease system in our sensor to express a range of colors, from green to yellow to red, according to the nitrate concentration in aquatic environments. When nitrate concentration is low, the sensor predominantly expresses *vsfGFP* and glows green; when the nitrate concentration is high, the sensor degrades *vsfGFP* through Lon protease system while simultaneously expressing *mScarlet*, making the sensor glow red. In the case of intermediate concentrations of nitrate, the sensor glows yellow from both *vsfGFP* and *mScarlet* expression.

The strength of the sensor is that it uses two different color mechanisms to evaluate the nitrate concentration. Unlike existing sensors that usually employ a one-color mechanism, it measures the concentration with two distinct colors, red and green. Such allows for more accurate measurement of nitrate concentration, as one mechanism can supplement the drawback of the other, which results in more stable and reliable data. Moreover, the construction principles for this nitrate biosensor lay the essential groundwork for designing biosensors for other chemical ions, such as chlorine and phosphorus. Additionally, the biosensor's affordable nature allows the local community to deploy an extensive surveillance network alongside the bodies of water to enable the monitoring of nitrate levels. This advantage encourages the local community to deal with eutrophication more proactively rather than relying on reactive measures, such as manual inspections. By collecting critical data on nitrate influx points, the biosensor also provides a foundation for implementing artificial intelligence and databases to predict and systematically prevent eutrophication.

Materials and Methods

Chemicals and materials

Sodium pentacyanonitrosylferrate(III) dihydrate (Alfa Aesar, A15656) and sodium nitrate (Alfa Aesar, A18668) were purchased from Thermofisher Scientific. Sodium alginate (SIGMA, A2158) and sodium chloride (SIGMA, S9888) were purchased from Merck.

Strains and Plasmids

The strains and plasmids utilized in the experiment are detailed in Table 1. DNA fragments, except for plasmid DNA, were synthesized by Integrated DNA Technologies (IDT) and Twist Bioscience. The synthesized DNA fragments were initially cloned into the Allinone T-vector (Biofact) and confirmed via Sanger sequencing. Subsequently, the verified DNA fragments were subcloned into the pET21b(+) or pSEVA211 plasmids.

Table 1. Strains and Plasmid

Strains	
DH5alpha	F- Φ 80lacZ Δ M15 Δ (lacZYA-argF)U169 deoR nupG recA1 endA1 hsdR1 7(rK- mK+) phoA glnV44 (supE44) thi-1 gyrA96 relA1, λ -
BW25141	F- λ - Δ (araB-araD)567 Δ lacZ4787(::rrnB3) Δ (phoB-phoR)580 galU95 Δ uidA3::pir+ recA1 Δ endA9::FRT rph-1 Δ (rhaB-rhaD)568 hsdR514
Plasmid	
pKJW01	bla P43::creRBS-mScarlet oriBR322
pKJW02	bla Pbla::RBS-mScarlet oriBR322
pKJW03	bla Pbla-mScarlet oriBR322
pKJW04	bla PlacI-mScarlet oriBR322
pKJW05	bla PlacIq1s-mScarlet oriBR322
pKJW06	bla PlacIq1-mScarlet oriBR322
pKJW07	bla P43::creRBS-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW08	bla Pbla::RBS-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW09	bla PlacI-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW10	bla PlacIq1-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW11	bla PlacIq1s-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW12	bla Pbla-nsrR;PyeaR-mScarlet-I3 oriBR322
pKJW13	bla Pbla::RBS-nsrR;PyeaR::RBS-mScarlet-I3 oriBR322
pKJW14	bla P43::creRBS-nsrR;PyeaR::RBS-mScarlet-I3 oriBR322
pKJW15	kan PL::RBS-vsfgFP-0::lva_tag oriR6K
pKJW16	kan PlacIq1CuO::RBS-vsfgFP-0::lva_tag oriR6K
pNitrateSafeGreen	kan PJEx-D::RBS-vsfgFP-0::lva_tag oriR6K
pNitrateCautionRed	bla Pbla::RBS-nsrR;PyeaR::RBS-mScarlet0I3::Mf_lon::EilR oriBR322

Construction of plasmids

To construct pNitrateSafeGreen, the region containing the R6K gamma replication origin and the chloramphenicol resistance marker was amplified by polymerase chain reaction (PCR) from the pSEVA211 plasmid using the following primer set: Forward, 5'-GATGGATCCCCAGGGGTCCCCAATAATTACG -3'; Reverse, 5'-GATGCATGCCCTAGA CAGCTGGGCGCGCC -3'. The amplified DNA was digested with SphI and BamHI and subsequently ligated using T4 DNA ligase into the PJEx-D-vsfgFP-0 plasmid. The plasmid used to determine the expression unit was constructed by excising the ORF of mScarlet with NdeI and XhoI and ligating it into pET21b(+). The promoter was digested with BglII and NdeI and

inserted into the reporter plasmid (pKJW01~pKJW06). For the nitrate sensing reporter construct, the promoter-reporter constructs were digested with NdeI and XhoI to remove the mScarlet gene. The nitrate sensing reporter unit (NsrR; PyeaR-mScarlet-I3 or NsrR; PyeaR::RBS-mScarlet-I3) was ligated into these sites, creating the plasmids pKJW07 through pKJW14. The pKJW13 plasmid was digested with NcoI and XhoI. The ORFs of Mf-Lon and EilR, which reduce the intracellular amount of LVA-tagged vsfGFP-0, were inserted into these restriction sites, creating the pNitrateCautionRed plasmid.

Determination of Expression Unit

Plasmid DNA in which the ORF of mScarlet is transcriptionally fused to various expression units (promoter and ribosome binding sites) was transformed into *E. coli*. Fluorescence was observed, and images were obtained using a fluorescence microscope equipped with a CCD camera on a colony grown appropriately on an LB agar plate. The obtained image was analyzed for pixel intensity by extracting only red channels using ImageJ. The pixel intensity value of fluorescence was expressed as a graph together with statistical analysis using Excel and Graphpad Prism programs.

Optimization of Reporter expression

Bacteria transformed with the reporter plasmid containing vsfGFP-0 and mScarlet-I3 were grown overnight in LB, dispensed 1 ml into a 1.5 ml microtube, centrifuged for 1 minute at 13,000 rpm, and observed for fluorescence. The pellet was diluted 1/100 in LB medium and transferred to a 24 well plate, which was then incubated for 2 hours, and treated with SNP or NaNO₃ as a nitrate source. The intensity of fluorescence at OD600 was measured using a multi-plate reader after the indicated time after nitrate source treatment.

Quantitative real-time polymerase chain reaction (RT-qPCR)

Total RNA was extracted from *E. coli* using the RNeasy mini kit (QIAGEN). Complementary DNA was prepared using the PrimeScript RT Reagent Kit. RT-qPCR was performed using TOPreal qPCR 2x PreMIX (SYBR Green with low ROX) on the QuantStudio™ 3 Real-Time PCR system. vsfGFP-0 was amplified with the following primers: Forward: 5'-CACAGGTGCAACTTGTGGAG-3', Reverse: CTTCATAGCT TGAGCGATCG. The delta Ct value was calculated using the Ct value of the internal control, 16s rRNA (Forward: GTTAATACCTTTGCTCATTGA, Reverse: ACCAGGGTATCTAAT CCTGTT). Finally, the fold change was calculated based on the delta Ct value of *E. coli* without pNitrateSafeGreen.

Preparation of Alginate bead

In the single-layer alginate bead, the transformed bacteria were grown in 250 ml of LB until the OD600 reached 0.8 and centrifuged at 4,000 rpm for 15 minutes to remove the supernatant. The cell was resuspended with LB so that the cell became 6.3×10^8 cells/ml. 500ul of the 2% sodium alginate solution and 500ul of the cell suspension were mixed well, dropped one

drop at a time into the 100 mM CaCl₂ solution by using a micropipette, to be recovered after 3 minutes and rinsed with DIW. The made alginate bead was immersed in PBS (pH 7.4) and stored at 4°C until used. A double-layered alginate bead was coated by rinsing the single-layered bead with DIW and immersing it in a sterilized 2% sodium alginate solution. The beads were then immersed in 100 mM CaCl₂ for 5 minutes, recovered, and rinsed with DIW, soaked in PBS, and refrigerated until use.

Results

To create a bicolored nitrogen detection system, we took advantage of NsrR-PyeaR system, a well-known nitrogen-responsive promoter system (Stern, A. M., and Zhu, J., 2014). Mechanistically, NsrR binds to PyeaR and inhibits transcription when the nitrogen level is low, while it dissociates from PyeaR when the nitrogen level is high, leading to PyeaR activation. Utilizing this regulatory mechanism, we designed constructs so that NsrR is under a constitutive promoter and mScarlet is under a PyeaR promoter (Figure 1). Under this design, only basal level of fluorescence should be detected when nitrate level is low since NsrR would fully inhibit PyeaR and therefore mScarlet expression, while fluorescence should gradually increase as nitrate level increases due to dissociation of NsrR from PyeaR promoter.

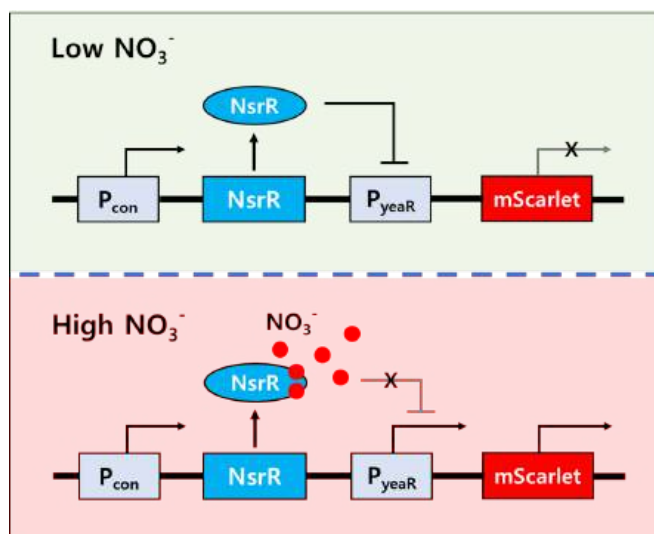


Figure 1. Essential Components for nitrate-sensing bio-sensor. The figure depicts the essential components required for nitrate sensing in the design of our bio-sensor. NsrR is constitutively expressed by the constitutive promoter (Pcon) and acts as a negative regulator of PyeaR. (Top) When the nitrate concentration is low, NsrR fully inhibits PyeaR and results in low levels of mScarlet expression. (Bottom) When nitrate concentration is high, it inhibits NsrR activity, which in turn promotes mScarlet expression.

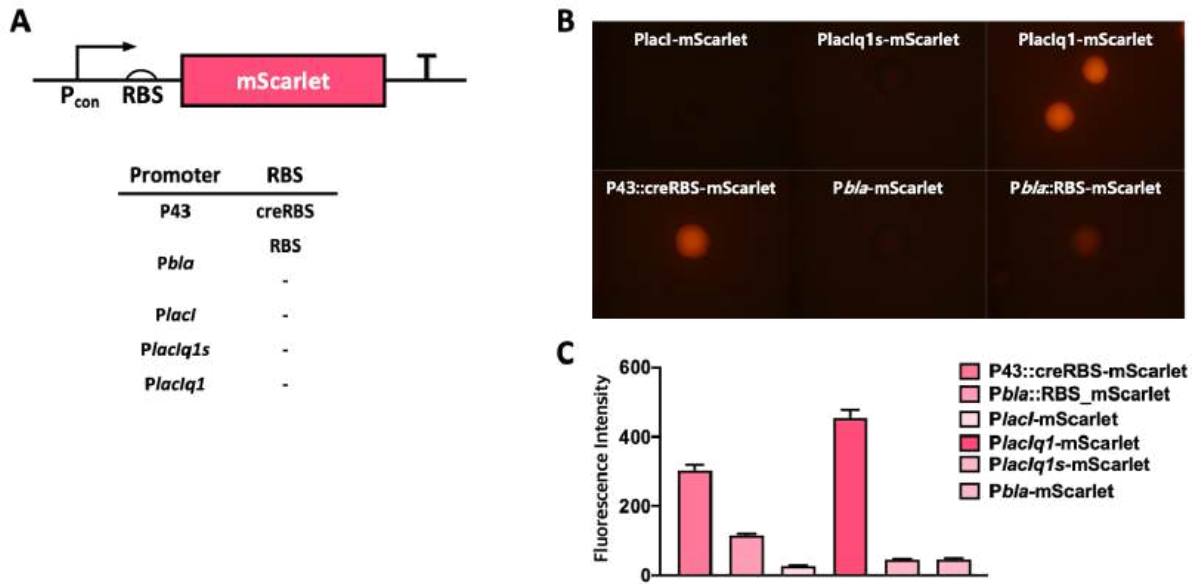


Figure 2. Measuring promoter intensity for expression of nsrR

(A) List of promoters in combination with RBS that were selected as candidates for the expression of mScarlet. (B) Images of *E. coli* expressing the constructs shown in (A) taken from a fluorescence microscope to visually check for mScarlet expression. (C) Quantification of mScarlet expression from images shown in (B). The quantification was achieved by measuring the pixel intensities from the appropriate channel (red) in ImageJ. n=10 (number of colonies analyzed)

To determine which of the constitutive promoters will be most suitable for NsrR expression, we tested six different expression units, using fluorescence intensity of mScarlet as our readout (Figure 2A). From our fluorescence readout assay, we classified *lacI* promoter and P43::creRBS as “strong” expression units, Pbla::RBS as “moderate,” and the remaining three units as “weak” expression units (Figure 2B, C).

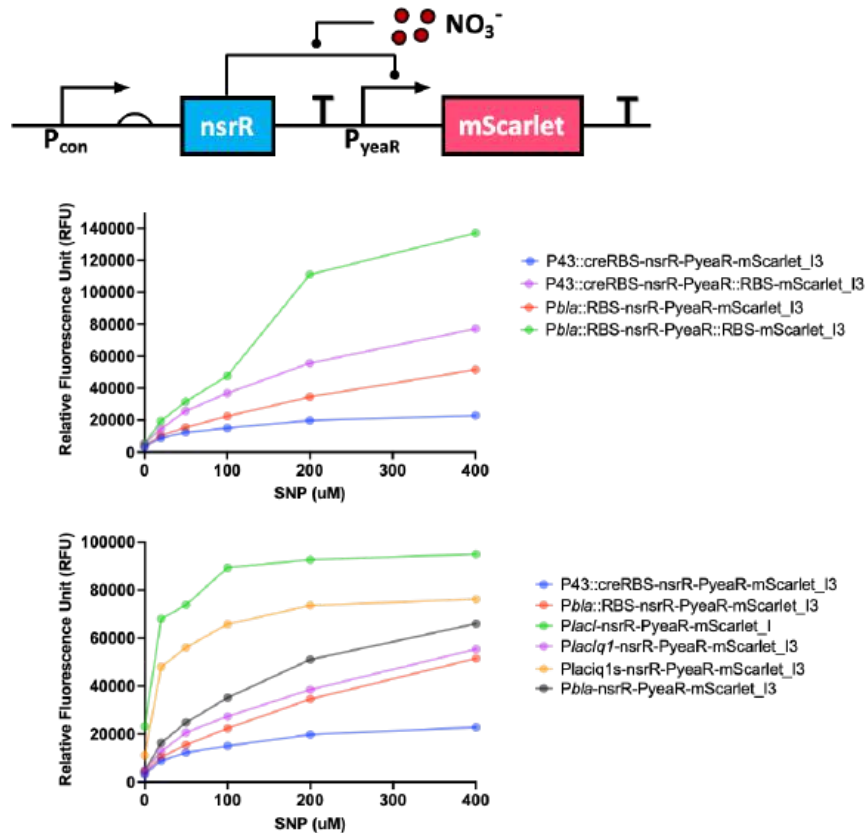


Figure 3. Selecting expression units for NO sensing

- (A) Plasmid construct design showing how negative regulation of *PyeaR* is achieved through expression of *NsrR*.
 (B) Comparison of nitrate sensing efficiency from different constructs with different intracellular *NsrR* expression levels. (n=3).

Next, we tested the sensitivity of our system following treatment with a nitrogen source. We used SNP at six different concentrations as nitrogen donors (Figure 3). Under two of the three “weak” promoters, *PlacI* and *PlacIq1s*, high fluorescence was detected even without the presence of SNP, indicating that *NsrR* expression was not high enough to fully inhibit *PyeaR*. On the other hand, in the case of *P43-creRBS*, which is a “strong” promoter, the fluorescence level remained at the basal level. Interestingly, however, the *P43-creRBS* system showed reduced sensitivity to nitrate, indicating that a high threshold concentration of nitrate is required to activate the nitrate sensor. From the three remaining expression units (*Pbla*, *Pbla-RBS*, and *PlacIq1*), a pattern of low basal levels at the lowest SNP concentration and a linear increase in fluorescence values with increasing SNP concentration was observed (Figure 3). Among the three expression units, we selected *Pbla-RBS* as the final candidate, which showed the most linear and noticeable increase in fluorescence with increasing SNP concentration.

Since the longevity and efficacy of our bio-sensor also depend on the viability and growth of *E. coli* that harbors our construct, we tested how SNP concentration affects bacterial growth. Our growth inhibition assay showed that increasing SNP concentration resulted in a

decreased *E. coli* growth rate and consequently a decrease in fluorescence (Figure 4A). To improve the viability of our sensor, we tested if growing cells in nutrient-rich BHI media improved the growth rate and fluorescence of *E. coli*. Indeed, *E. coli* grown in a BHI medium showed much higher resilience towards increasing SNP concentration, which was evident in both the growth inhibition assay and fluorescence intensity measurements (Figure 4). Therefore, we concluded that our biosensor should be grown in a nutrient-rich environment for maximal efficacy.

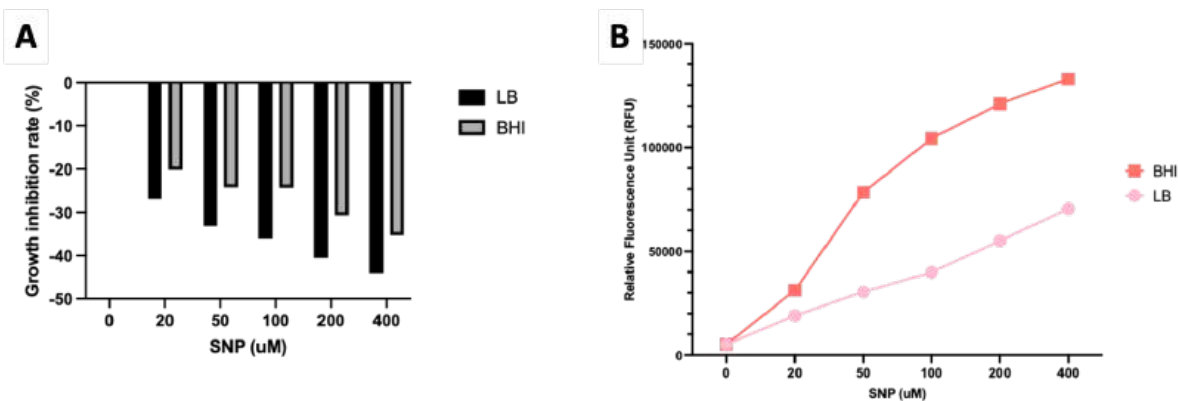


Figure 4. Optimizing NO biosensor culture conditions

(A) Change in bacterial growth rate depending on the Culture medium. OD600 was measured after 4 hours of different doses of SNP treatment on *E. coli* grown in either BHI or LB medium. (B) Quantification of mScarlet fluorescence intensity under the same conditions described in (A). The shown graph is a representative of n=3 biological replicates.

Our next step was to design and incorporate a separate construct expressing GFP so that our sensor emits green color under low nitrate conditions while it emits red under high nitrate conditions. We created three different constructs under three different promoters that were combined into the vsfGFP-0 coding sequence and tested them individually for vsfGFP expression using fluorescence measurement. Among the three promoter systems tested, the PJex-D construct displayed the strongest expression of vsfGFP under normal conditions (Figure 5). To most efficiently inhibit vsfGFP expression with increasing nitrate concentration, we also additionally included EilR, a repressor that inhibits PJex-D, into our design. EilR was cloned into the PyeaR module so that activation of PyeaR due to high nitrate level would result in EilR expression and subsequently decrease GFP expression from PJex-D (Figure 6A).

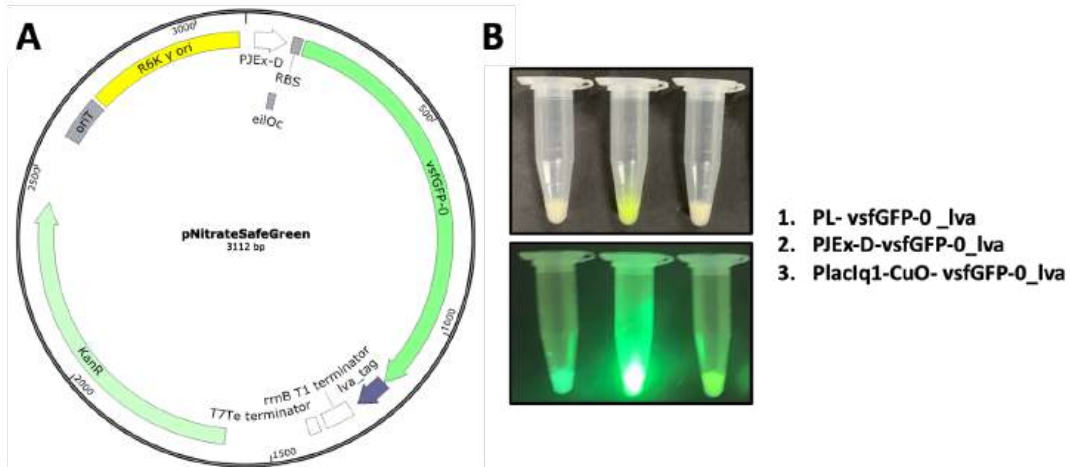


Figure 5. Expression of GFP in the constructed plasmid, pNitratesafeGreen
 (A) Plasmid map of pNitratesafeGreen. (B) Images of tubes containing *E. coli* expressing the reporter constructed are depicted in (A). Visualization was achieved by shining royal blue LED light onto the tubes.

To test the inhibition efficiency of PJEx-D by EilR, we added 400 μ M SNP to *E. coli* that contained our constructs, which was then lysed and subjected to RNA extraction and RT-qPCR. After SNP treatment, the amount of vsfGFP fluorescence gradually decreased, and after 4 hours, the transcription of vsfGFP dropped to the basal level (Figure 6A). This indicated that the P_{yeaR}-EilR system could robustly inhibit vsfGFP expression from PJEx-D.

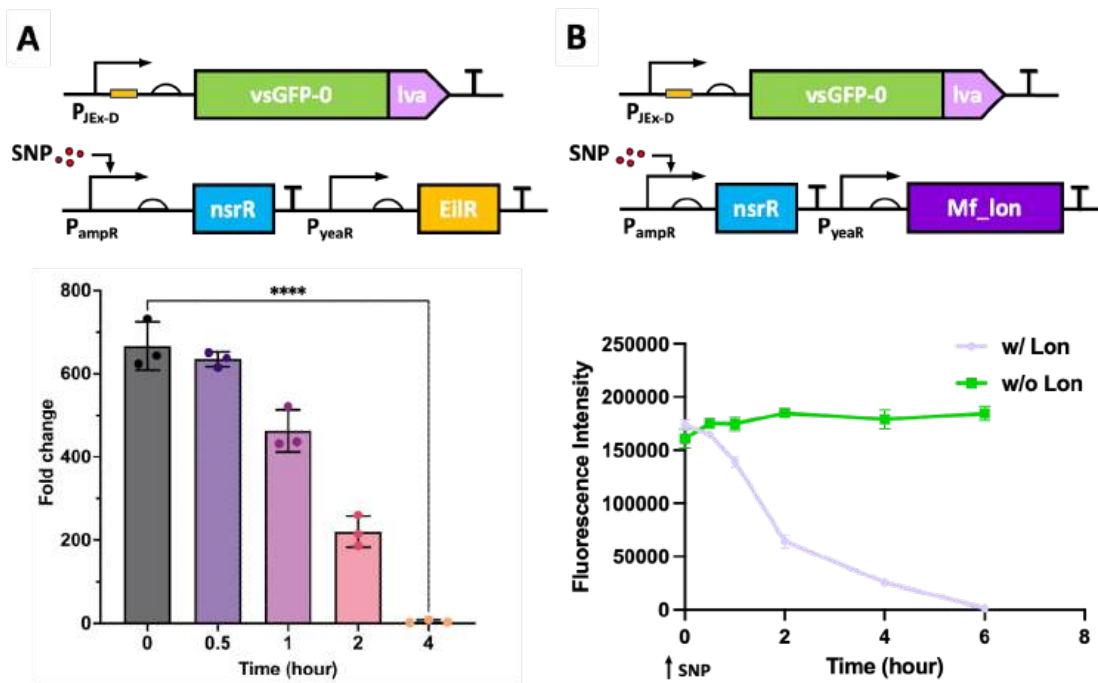


Figure 6. Reduction of intracellular vsfGFP through regulation at the mRNA and protein levels

(A) (Top) Simplified diagram of our designed construct for testing vsGFP transcript levels. (Bottom) The fold change in vsGFP transcript level was quantified using RT-qPCR following the addition of 400uM SNP. (n=3, p < 0.001) (B) (Top) Simplified diagram of our designed construct for testing vsGFP expression levels at the protein level. (Bottom) Quantification of vsGFP fluorescence intensity over time following SNP treatment in vsGFP constructs containing Lon tag and vsGFP not containing Lon tag.

While inhibition of PJex-D through EilR was successful, the fact that it took almost 4 hours for vsfGFP expression to drop to the basal level posed a practical concern since detection of fluctuating nitrate levels should ideally be much faster. To even further improve our system, we also tested if we could incorporate a degron into vsfGFP so that vsfGFP can be actively degraded while its transcription is simultaneously inhibited by EilR. To achieve this, we tagged the vsfGFP c-terminal region with an LVA tag, which could be recognized and degraded specifically by Lon protease. Similar to the previous experiment with EilR, we co-expressed two constructs, one expressing vsfGFP under PJex-D and the other expressing Lon under PyeaR (Figure 6B). As a control, we also tested cells that only expressed vsfGFP under PJex-D. Compared to the control cells that only expressed vsfGFP-LVA, cells that expressed both Lon and vsfGFP-LVA showed a significantly faster decrease in fluorescence after the addition of SNP, in which the fluorescence dropped to the basal level around the 2-hour mark (Figure 6B).

With all the results from the aforementioned experiments, we were able to construct two constructs, pNitrateSafeGreen (pNSG) and pNitrateCautionRed (pNCR). pNSG contains vsfGFP-LVA under PJex-D, while pNCR contains NsrR under PampR and mScarlet-Lon-EilR under PyeaR (Figure 7A). To verify that the fluorescence emitted by the biosensor shifts dynamically as it is exposed to SNP, *E. coli* with both plasmids were exposed to 400 μ M of SNP, and the change in fluorescence intensity was measured. Following SNP treatment, red fluorescence was visible even with bare eyes 2 hours after exposing it to SNP. The fluorescence intensity measurements corroborated this observation, where the addition of SNP caused green fluorescence to exponentially decrease, while red fluorescence increased gradually over time (Figure 7B, C). This final design of our biosensor, which is *E. coli* transformed with both pNSG and pNCR, was named “Monitro.”

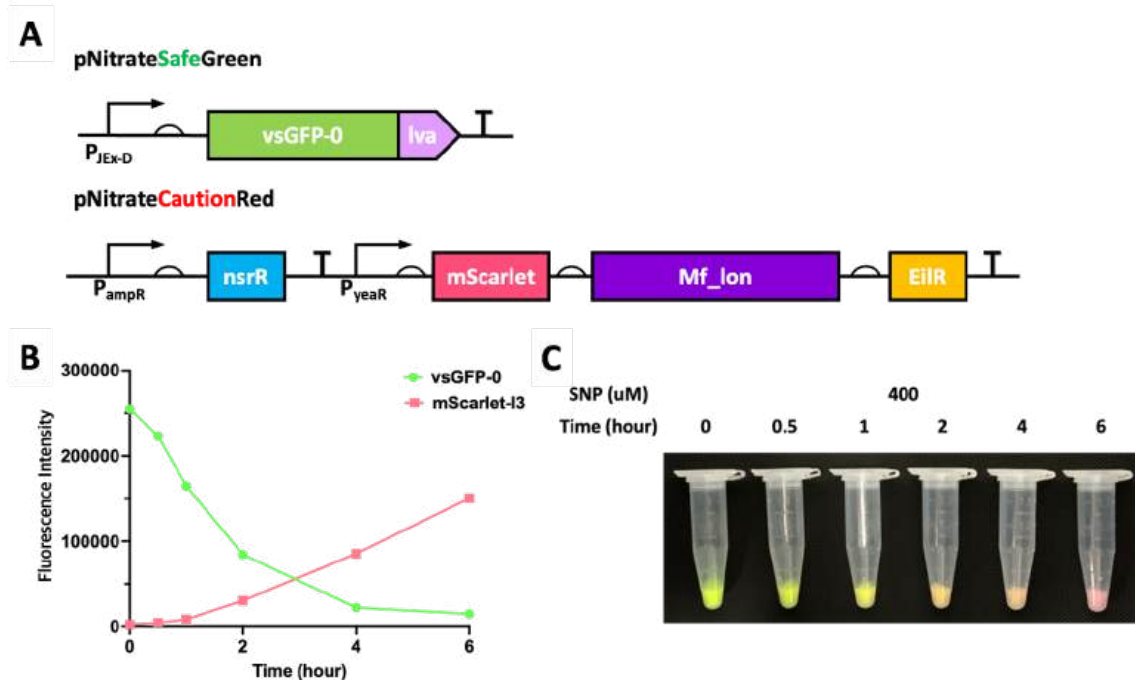


Figure 7. Monitro, a biosensor that detects NO

(A) Schematic of a biosensor that changes color depending on the concentration of nitrate. (B) Changes in fluorescence intensity of vsGFP (green) and mScarlet (red) over time following SNP treatment (400 μ M) on *E. coli* co-expressing pNitrateSafeGreen and pNitrateCautionRed. (C) Representative image showing the visible color change from green to red following SNP treatment described in (B).

For practical usage, Monitro (OD600 = 0.8) was packaged in a double-layered alginate bead to apply the biosensor in real-world conditions (Figure 8A). The bead was double-layered to prevent leakage that was observed in single-layered beads. Then, the bead was exposed to 10 mM of NaNO₃, which visibly turned red after 6 hours (Figure 8B). Additionally, an RGB sensor was employed to quantitatively measure the color, which corroborated our observation (Figure 8C).

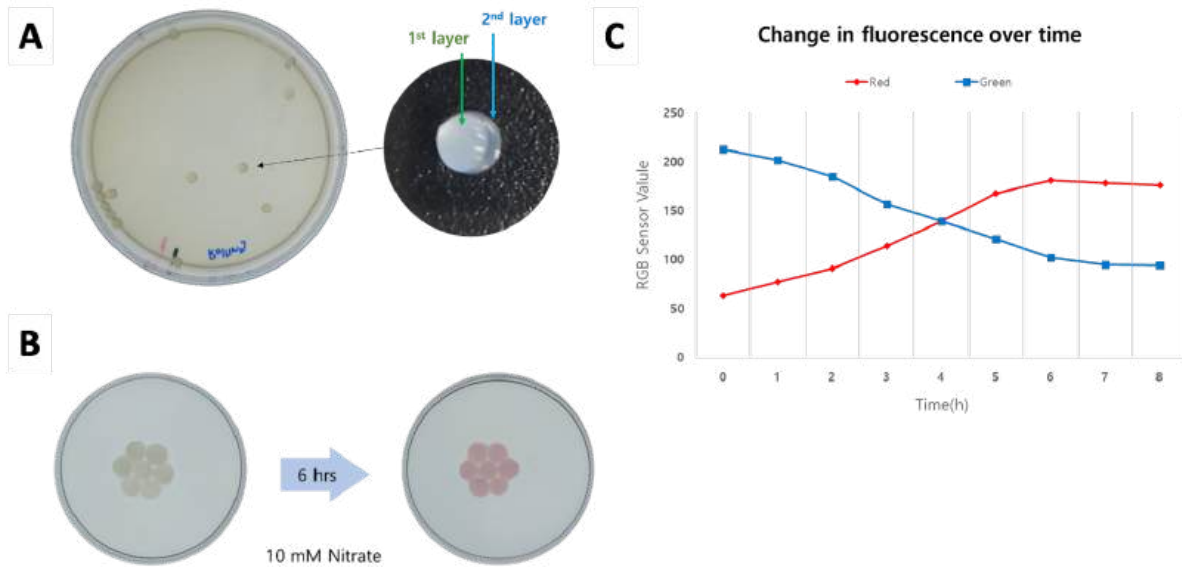


Figure 8. Color change of a double-layered nitrate biosensor (A) Representative image of a double-layered nitrate biosensor (B) Color change of a double-layered nitrate biosensor after 6 hours of 10mM nitrate treatment (C) Quantification of change in fluorescence over time of double-layered nitrate biosensor shown in (A) and (B).

Discussion

In this study, we successfully employed red and green fluorescence to visually express the nitrate levels in *E. coli* encapsulated in double-layered alginate beads and demonstrated the applicability in the real world. The existing colorimetric nitrate sensors often use one type of color, so they are limited to expressing the mere presence of nitrates. On the contrary, the three-tiered color scale (red, yellow, green) enabled by red and green fluorescence allowed for more sophisticated measurements of nitrate. Monitro comprises two plasmids: pNitrateSafeGreen and pNitrateCautionRed. As the name suggests, pNitrateSafeGreen emits green fluorescence in normal conditions. In contrast, pNitrateCautionRed reduces the green and expresses red fluorescence under the presence of nitrate. More specifically, pNitrateSafeGreen contains a coding sequence of vsfGFP that is controlled by the negative inducible promoter (JEx-D promoter). JEx-D promoter contains a binding site EilOc for EilR, which acts as a transcription repressor. If EilR concentration increases within the *E. coli*, they, in turn, bind to EilOc, which inactivates the JEx-D promoter, terminating the creation of green fluorescence. Additionally, the vsfGFP has an LVA tag fused to the c-terminus. Lon Protease recognizes the proteins with LVA tags and degrades them, reducing the green fluorescence.

The mentioned two regulatory loops are integral processes in converting green to red amidst the presence of nitrate. The sensitivity of the biosensor to nitrate varies depending on the concentration of NsrR in *E. coli*, which controls the activation of the yearR promoter negatively. If NsrR is largely produced by a strong promoter, the change in fluorescence intensity is not observed in low concentrations. On the other hand, if NsrR is maintained at a low concentration, it saturates at a low nitrate concentration. If NsrR is expressed by a weak expression unit, the

yeaR promoter does not close completely, and nitrate sensitivity increases, so NsrR cannot control the yearR promoter at 100 μM SNP, limiting its functioning range. To allow Monitro to function in a broad range, a moderate expression unit was selected, leading to a confirmed working range of 0 to 400 μM . The accepted amount of NO_x in water set by WHO is at least 10 mg/L (\approx 160 μM), which is within the detecting limits of Monitro. Furthermore, Monitro was packaged in an alginate bead to allow for real-time monitoring in versatile conditions. Since the single-layered alginate bead was insufficient in safely containing *E. coli*, which undermines the sensor's performance and poses a detrimental threat to the aquatic environment, another layer was added to remedy this issue.

In summary, Monitro employed fluorescent protein as an indicator to gauge the contamination of NO. Moreover, the cost was greatly reduced by using an RGB sensor and Arduino instead of relying on a fluorescence sensor to quantify the color change of alginate beads. We are currently in the step of testing a new expression unit to create a biosensor specialized for detecting extremely high or low concentrations of NO by fine-tuning the amount of NsrR. We are also verifying whether the freeze-drying of the Monitro packaged in a double-layered alginate bead can increase the longevity of the sensor without impairing its performance.

Works Cited

- Almeida, M. G., Serra, A., Silveira, C. M., & Moura, J. G. (2010). Nitrite Biosensing via Selective Enzymes—A Long but Promising Route. *Sensors (Basel, Switzerland)*, 10(12), 11530-11555. <https://doi.org/10.3390/s101211530>
- Cameron, D. E., & Collins, J. J. (2014). Tunable protein degradation in bacteria. *Nature biotechnology*, 32(12), 1276-1281. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3231041/>
- Chislock, M. F., Doster, E., Zitomer, R. A. & Wilson, A. E. (2013) Eutrophication: Causes, Consequences, and Controls in Aquatic Ecosystems. *Nature Education Knowledge* 4(4):10
- NOAA. (2017) What is eutrophication? National Ocean Service of National Oceanic and Atmospheric Administration (US). <https://oceanservice.noaa.gov/facts/eutrophication.html>
- Nakdong River Network (2022) Yeongsan River rice, microcystin detected. Korean Federation for Environmental Movement in Pusan (South Korea), <https://pusan.kfem.or.kr/data/board/act03/files/74d17bd0dceb03a2cddc2167009d2d5d.pdf>
- Plumeré N. (2013). Interferences from oxygen reduction reactions in bioelectroanalytical measurements: the case study of nitrate and nitrite biosensors. *Analytical and bioanalytical chemistry*, 405(11), 3731–3738. <https://doi.org/10.1007/s00216-013-6827-z>
- Ruegg, T. L., Pereira, J. H., Chen, J. C., DeGiovanni, A., Novichkov, P., Mutalik, V. K., Tomaleri, G. P., Singer, S. W., Hillson, N. J., & Simmons, B. A. (2018). Jungle Express is a versatile repressor system for tight transcriptional control. *Nature communications*,9(1), 3617.
- Seo, S.-O., & Schmidt-Dannert, C. (2019). Development of a synthetic cumate-inducible gene expression system for *Bacillus*. *Applied microbiology and biotechnology*, 103, 303-313.
- Smith, Val H. "Eutrophication of Freshwater and Coastal Marine Ecosystems: A Global Problem." *BioScience*, vol. 51, no. 3, 2001, pp. 227-234. Oxford University Press, <https://academic.oup.com/bioscience/article/51/3/227/256199>.
- Stern, Andrew M., and Jun Zhu. "An introduction to nitric oxide sensing and response in bacteria." *Advances in applied microbiology* 87 (2014): 187-220.
- Valdez-Cruz, N. A., Caspeta, L., Pérez, N. O., Ramírez, O. T., & Trujillo-Roldán, M. A.

Tourism Dynamics and Economic Growth in Norway: A Dual-Perspective Survey

Analysis By Maanav Chittireddy

Abstract

The Travel & Tourism sector is a significant component of the global economy, contributing substantially to GDP and employment. Norway, known for its natural beauty, relies heavily on tourism as an economic pillar. This paper evaluates the economic contributions of tourism to the Norwegian economy, explores tourist spending behaviors, and assesses the perceptions of local residents on tourism's impact. The study provides insights that can guide sustainable policy-making and development in the tourism sector.

1 Introduction

In 2023, the Travel & Tourism sector composed 9.1% of the global GDP and 313 million jobs [1] [2]. From food costs to lodging fees, the travel and tourism industry provides a plethora of opportunities to stimulate economies, generate revenue, and create jobs. For nations that are reliant on their terrestrial landscapes and natural beauty, tourism is a key pillar of their economy. Norway, the country home to most fjords in the world, generated 541.5 billion Norwegian kroner in 2022 from the tourism industry alone [3]. Cities and attractions such as Oslo, Bergen, Geirangerfjord, and Tromsø are among the most popular places for tourists to visit.

Due to the variety of backgrounds tourists come from, the definition of tourism may be ambiguous. Tourists are most often defined as people who travel to an area because of vacations, as well as people who travel to an area because of work, etc. [8] Note that this definition of a tourist is activity-based and purpose-driven.

Despite the tourism industry's clear economic benefits, the tourists' overall impact is multifaceted. Although tourism does improve infrastructure and the circular flow of money in an economy, the increased population of foreigners could exert a strain on resources, decrease living space, establish social tensions, and raise living costs. Understanding the holistic impact of tourism is key to efficient policy-making and sustainable development.

The objectives of this study are to (1) evaluate the economic contributions of tourism, (2) understand tourist spending behaviors and satisfaction, and (3) gauge residents' perceptions of tourism's impact. This study aims to holistically analyze the economic impact of tourists by considering both the tourists themselves and the residents who are affected by their presence.

With the results of this study, local governments, tourism boards, and businesses could learn more about the benefits and challenges associated with tourism. Insights from this research can help develop and modify initiatives to maximize the positive impacts of tourism while mitigating its negative consequences. The scope of this study encompasses major Norwegian attractions over the past five years, with a focus on both economic indicators and quantitative and qualitative perceptions from key stakeholders. In the following sections, this paper will outline the methodology used for data collection and analysis, discuss the results, and provide a comprehensive conclusion based on the findings.

2 Methodology

2.1 Research Design

This study aims to be holistic in its data collection approach, considering both qualitative and quantitative data. The primary research questions are as follows:

1. To what extent does tourism impact the Norwegian economy?
2. What are the spending patterns among tourists, and how satisfied are they?
3. How do Norwegian residents perceive the impact of tourism?

Surveys were designed for primary data collection to address these questions. Pilot testing was conducted to refine the survey questions, optimizing clarity and relevance.

2.2 Tourist Survey

The tourist survey was designed to gauge tourist satisfaction, spending habits, and the recreational activities they engage in. It included both quantitative and qualitative questions, with tourists responding to multiple-choice and open-ended items:

2.2.1 Quantitative Questions

1. How much did you spend on accommodation during your stay (USD)?
 - <\$50
 - \$50-\$100
 - \$100-\$200
 - >\$200
2. How many days are you staying in Norway?
3. Which activities did you spend the most money on?
 - Sightseeing
 - Dining
 - Shopping
 - Outdoor Activities
 - Other
 -

2.2.2 Qualitative Questions

1. What did you enjoy most about your visit to Norway?
2. Do you have any suggestions for improving the tourist experience in Norway?

2.3 Local Resident Survey

The local resident survey aimed to assess the frequency of tourist interactions with local residents and how these residents perceived the social, economic, and environmental impacts of tourism on their communities:

2.3.1 Quantitative Questions

1. How often do you interact with tourists in your city?
 - Daily
 - Weekly
 - Monthly
 - Rarely
 - Never
2. Do you believe tourism positively impacts the local economy?
 - Strongly Agree
 - Agree
 - Neutral
 - Disagree
 - Strongly Disagree
3. Have you noticed changes in local job opportunities due to tourism?
 - Increased
 - Decreased
 - No Change
 - Not Sure

2.3.2 Qualitative Questions

1. What are the main challenges or negative impacts of tourism in your community?
2. How could the local government better manage tourism to benefit the community?

2.4 Data Collection and Processing

Quantitative survey responses were analyzed using percentages and distributions. Qualitative data were examined for recurring themes and patterns. Efforts were made to address potential biases through random sampling and pilot testing.

3 Results

3.1 Resident Survey Results

Interaction Frequency	Number of Residents
Daily	22
Weekly	12
Monthly	8
Rarely	5
Never	3

Table 1: Frequency of Residents' Interactions with Tourists

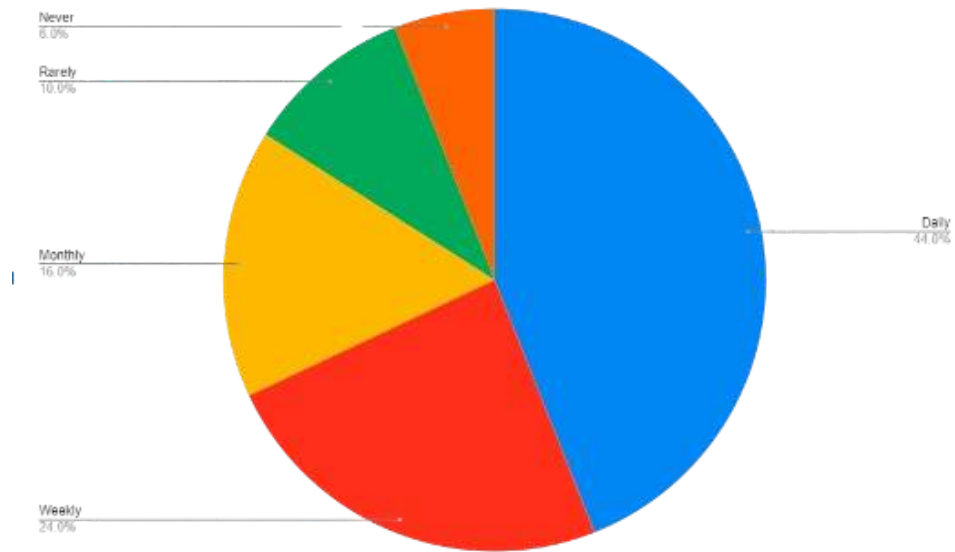


Figure 1: Frequency of Residents' Interactions with Tourists

Resident Perception	Number of Residents
Increased	10
Decreased	6
No Change	9
Not Sure	25

Table 2: Residents' perceptions on whether there has been an increase in job opportunities due to tourism

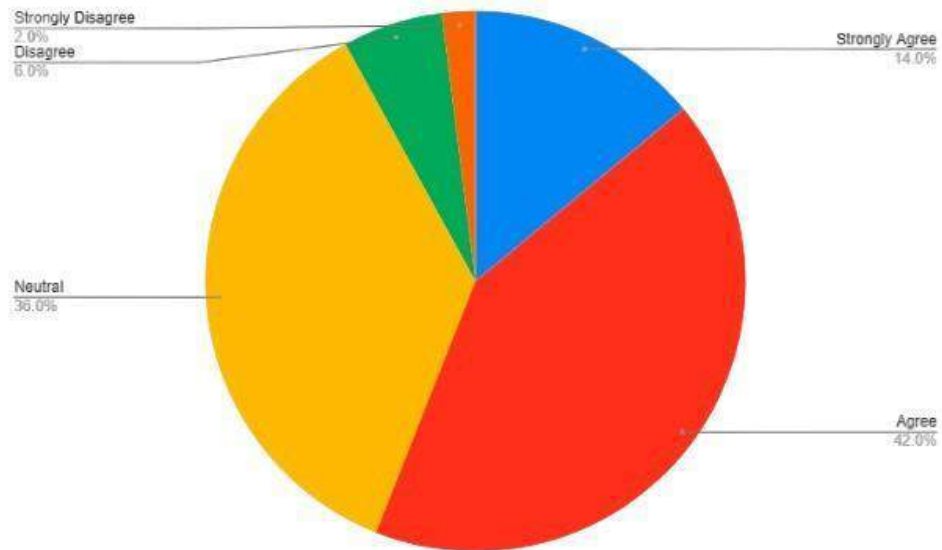


Figure 2: Residents' perceptions on whether there has been an increase in job opportunities due to tourism

Resident Perception	Number of Residents
Strongly Agree	7
Agree	21
Neutral	18
Disagree	3
Strongly Disagree	1

Table 3: Residents' opinion on whether tourism is positive for the economy

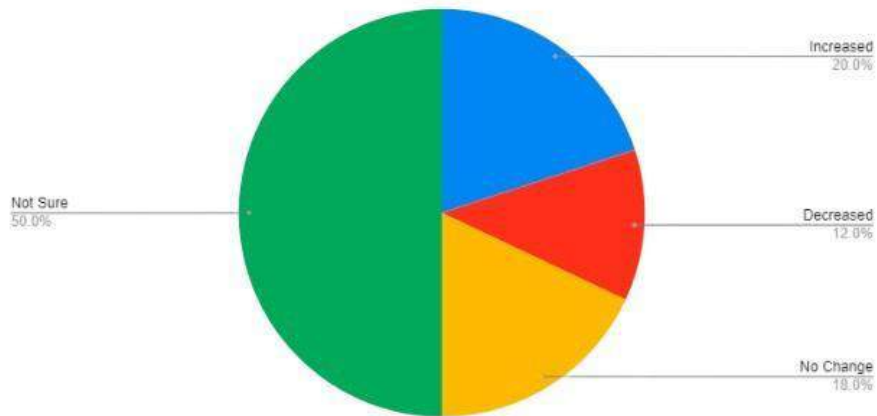


Figure 3: Residents' perceptions on whether there has been an increase in job opportunities due to tourism

3.2 Tourist Survey Results

Spending Amount (USD)	Number of Tourists
100-200	5
200-500	35
500-1000	7
1000+	3

Table 4: Tourist spending during stay in Norway

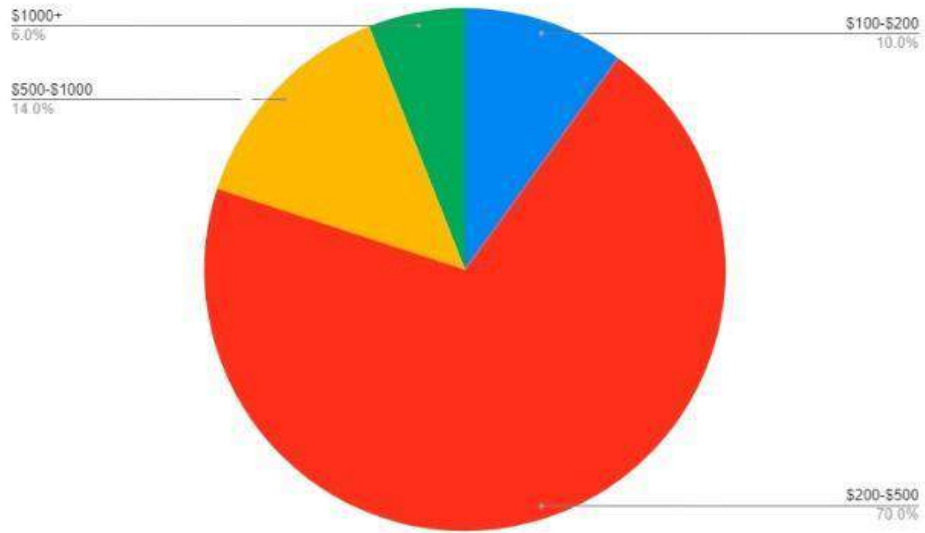


Figure 4: Tourist spending during stay in Norway

Recorded Answers (In days)	# Of Tourists
3	2
4	2
7	10
12	11
14	5
15	5
17	3
19	3
20	4
23	2
25	3

Table 5: Tourists' length of stay in Norway

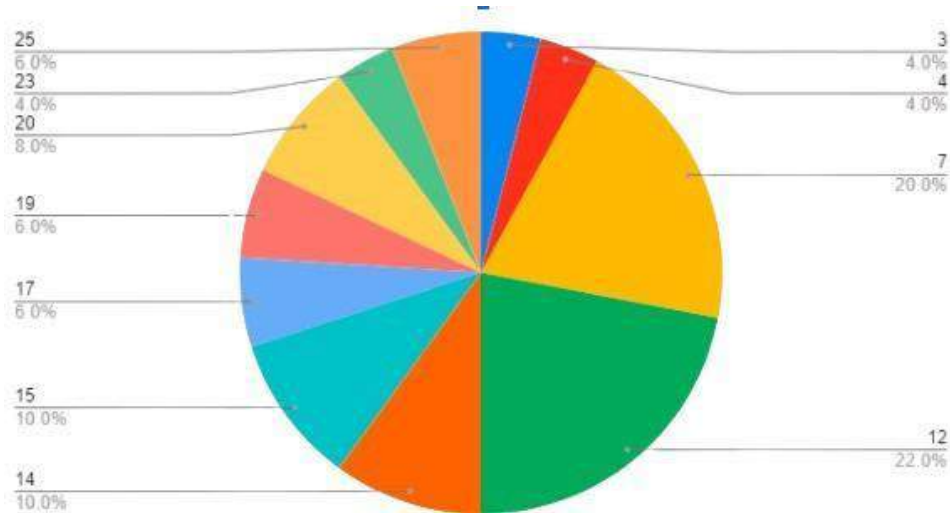


Figure 5: Tourists' length of stay in Norway

Activities	# Of Tourists
Hiking	4
Food and Lodging	31
Transportation	11
Fishing	1
Museums and National Parks	3

Table 6: Activities that tourists spent the most money on

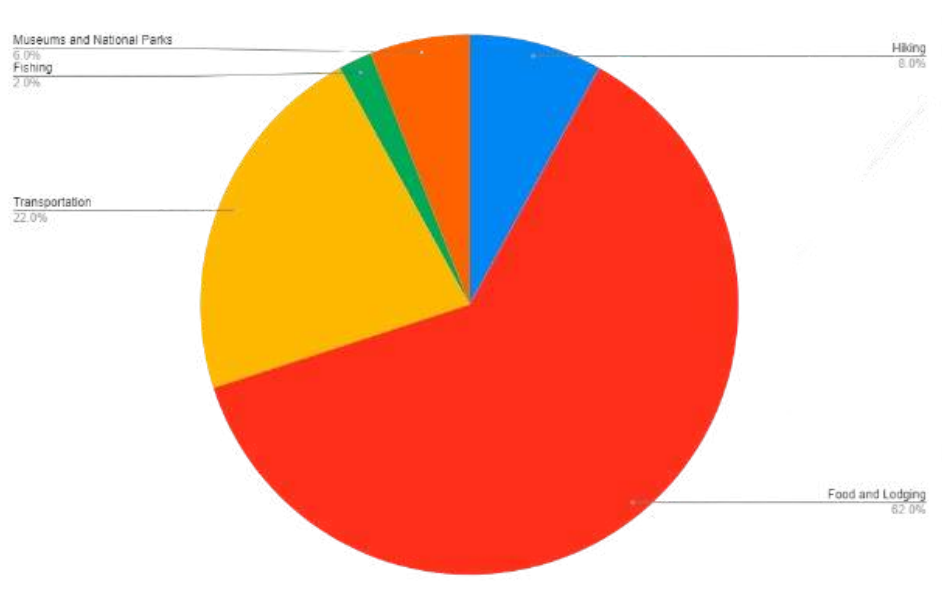


Figure 6: Activities that tourists spent the most money on

4 Discussion

As a country renowned for its natural beauty, Norway's tourism industry plays an integral role in its economy. Tourism contributes significantly to the Norwegian economy, accounting for 3.6% of GDP and 6.1% of employment [4]. This GDP contribution is reflected in Figure 4 , which shows that 70% of surveyed tourists spend \$200-\$500 during their stay in Norway. Although Norway benefits economically from tourism, this potential has not been fully optimized—tourism provides 6.1% of Norway's total jobs, yet 50% of residents have not noticed changes in local job opportunities due to tourism. This discrepancy suggests that while employment in the tourism sector is available, many Norwegians are unaware of these job opportunities. Promoting jobs in the tourism industry for residents can help mitigate the 4.1% unemployment rate and enhance the tourist experience by having more residents guide tourists [5].

When asked, “Do you have any suggestions for improving the tourist experience in Norway?”, one tourist responded, “Having more guidance on places to park, sights to see, or camping spots would be great!”. Encouraging residents to provide this guidance would have significant economic benefits. Specifically, establishing more information centers throughout Norway and having residents serve at those centers, with effective outreach about the employment opportunities, would be beneficial. With more residents involved in the tourism sector, they could play a more active role in how tourism affects their homeland. Of all the residents surveyed, 44.2% interact with tourists daily, and 56% either strongly agree or agree that tourism is beneficial for the economy. Nonetheless, this latter number could be further increased by improving residents' satisfaction with tourist behavior. Providing information on what to do or not do when touring Norway could ensure that all stakeholders are satisfied, as the residents' homeland is treated respectfully and tourists receive information that will improve the quality of their trip.

One goal that residents have emphasized is sustainability in the tourism sector. When asked, “What are the main challenges or negative impacts of tourism in your community?”, 52% of interviewed residents expressed concerns about the sustainability of tourism and preserving the natural beauty of their homeland. Additionally, seven residents focused on reducing greenhouse gas emissions when asked, “How could the local government better manage tourism to benefit the community?”. Although existing Norwegian tourism offers meet sustainability demands at a more or less satisfactory level, a major unresolved problem is the reduction of GHG emissions, primarily caused by the aviation industry and cruise line activities [6]. Addressing this concern by encouraging carbon offset programs with tax benefits and exploring Sustainable Aviation Fuels (SAFs) and Liquefied Natural Gases (LNG) for airlines and cruises, respectively, could reduce the 44% of residents who are neutral, disagree, or strongly disagree that tourism is beneficial for the economy.

When analyzing tourist satisfaction, it is essential to consider both length of stay (LOS) and spending patterns. A longer LOS provides more detailed information about the tourist experience in Norway and spending patterns, allowing for connections between spending and tourist

sentiment to be discerned. In this study, the majority of tourists stayed in Norway for 12-17 days, providing them with ample time to experience the country fully. Sixty-two percent of these tourists spent the most money on food and accommodations, which corresponds to the World Travel and Tourism Council's findings that food and accommodation account for the first and second highest tourist expenditures, respectively. Although tourists spent the most money on food and accommodations, the most complaints were received regarding the transportation situation, on which only 22% of tourists spent the most money. Several tourists interviewed, particularly those from the U.S., complained that at certain commercial centers, such as gas stations, they were required to enter a PIN to make a credit card transaction. However, in the U.S., this is not necessary, so these tourists were unaware of their PIN and were unable to purchase fuel at several gas stations they visited. Furthermore, 24% of tourists reported experiencing some form of property theft when asked, "Do you have any suggestions for improving the tourist experience in Norway?". This corresponds with 111,584 reported cases of property theft in the past year [7].

5 Conclusion

Although Norway's tourism sector plays a pivotal economic and social role, contributing significantly to GDP and employment, its potential in these areas has not been fully optimized. Environmental concerns about GHG emissions from tourism persist among Norwegian residents, tourists face crime and transaction difficulties, and many Norwegian residents are unaware of the employment opportunities available to them. Nonetheless, the emphasis that Norway places on tourism cannot be overlooked, with initiatives such as Co2rism and the National Tourism Strategy. Addressing the specific concerns mentioned would help improve the tourism experience for all stakeholders and position Norway at the forefront of the global tourism industry.

Works Cited

- [1]“News Article: World Travel & Tourism Council (WTTC).” World Travel & Tourism Council, World Travel & Tourism Council, 2 Feb. 2024, wttc.org/news-article/travel-and-tourism-set-to-break-all-records-in-2024-reveals-wttc.
- [2] NHO Reiseliv. Turismens Betydning for Norge. NHO Reiseliv.<https://www.nhoreiseliv.no/contentassets/c5f3c5015f7840e1bb0184f88b5a3be3/turismens-betydning-for-norge.pdf>.
- [3]Lopez, Ana M. ”Travel and Tourism: Total Contribution to GDP in Norway.” Statista, 2024. <https://www.statista.com/statistics/806905/travel-tourism-total-gdp-contribution-norway/>.
- [4]OECD. Norway. Semantic Scholar. <https://www.semanticscholar.org/paper/Norway/b38d1fef7600ccf62ded09377390b19108028aff>.
- [5]YCharts. ”Norway Unemployment Rate.” YCharts,2024.https://ycharts.com/indicators/norway_unemployment_rate.
- [6] Fayzullina, Svetlana. ”Evaluation of Sustainable Tourism in Norway.” Research-Gate, 2024. https://www.researchgate.net/publication/363113489_Evaluation_of_sustainable_tourism_in_Norway.
- [7] Statistics Norway. “Anmeldte Lovbrudd og Ofre.” Statistics Norway, 2024. <https://www.ssb.no/en/sosiale-forhold-og-kriminalitet/kriminalitet-og-rettsvesen/statistikk/anmeldte-lovbrudd-og-ofre>.
- [8] Lundberg, Donald E., Krishnamoorthy, Murthy, and Stavenga, Mike H. ”Tourism Economics.” *Annals of Tourism Research*, vol. 25, no. 1, 1998, pp. 135-153.<https://www.sciencedirect.com/science/article/abs/pii/S016073839800019X>.

The Familiar Melody of Memorization: A Musical Approach to Emily Dickinson **By Tessa Gumberg**

I have long personified the memorization of written material as a duplicitous character: simultaneously a disagreeable enemy and obliging companion. Internalizing entirely foreign, exterior content is a process rooted in the unfamiliar. Thus, the act of memorizing -- in its most elementary definition of transporting knowledge from an external format to the brain -- evokes a sentiment of discomfort. While it is an uncomfortable process itself to manifest something unfamiliar in the otherwise familiar context of my own brain, I have always found a comforting undertone in the resulting ability to effortlessly recall. Maintaining an internal arsenal of comprehension provides a sense of security; to know I can scour my archive at any time and usher a fully formed fact or phrase to the front, prepared to wield in discourse or deliberation, is simply reassuring. In an attempt to lasso the comforting facet of memorization and use it to combat the discomfort of the unfamiliar, I rooted my process of internalizing Emily Dickinson's poems 153 and 182 in what I consider the most familiar: the act of committing music to memory. As I mirrored my trusted approach of memorizing music in this poetic process, I discovered methods to compartmentalize Dickinson's flowing stanzas to the portion of my subconscious mind that sporadic Taylor Swift lyrics and commercial jingles inhabit. Simply put, I attempted to get the poems stuck in my head. Through this musically-influenced process, I not only focused upon the familiar, but additionally formed parallels between the widely analogous form of music to that of poetry, bolstering my interpretation of the text beyond my original analysis.

My recollection of music is a steadfast constant, accompanying me in the pursuit of drastically unrelated tasks. I often marvel at music's ability to manifest itself so profoundly in my mind as I hum the melodies of songs I memorized a full decade ago. Music retains an indelible quality unlike any written text or spoken word. With this notion, in addition to the bulk of my effective procedure with memorization pertaining to my career as a pianist, I began the process of internalizing poems 153 and 182 with the ultimate goal of Dickinson's phrases flowing as effortlessly as pitches from the ivory keys of a piano.

In the early process of learning a song, I consistently memorize the progression of the chords before the specific lyrical composition of each verse. As echoes of the song's melody plant themselves in the soil of my mind, they progressively burrow further into my subconscious until eventually, the song is "stuck" and I cannot cease humming the tune. This marks the inciting step of the musical memorization process and with conscious attention and practice, these buried seeds will blossom into fully formed blooms of song. I attempted to recreate this process by uncovering a poetic equivalent to a song's tune. I landed on assigning thematic titles -- a single word that captured the central focus -- to each of the poem's stanzas. For instance, in the second stanza of 153, Dickinson characterizes death as enigmatic, reinforcing its inevitable presence as one who exists yet "[n]ever was a [b]oy" (Dickinson 6). Considering its content, I mentally assigned this stanza the title of "backstory." Applying this technique of grouping and naming to each portion of the poem and memorizing that attributed sequence created an

equivalent form to the chord progression of a song. Rather than C Major, G Major, A Major, and subsequent resolution in F Major, poem 153 followed the stanzaic evolution of “opening,” “backstory,” “descriptions,” and finally, “bird cycle.” Once I could effortlessly recite this summary of events, I had essentially internalized the “tune” of the poem and was equipped to begin narrowing in on the intricate words of the stanzas themselves. Not only did this process of grouping further my ethos of familiarizing myself with the poem’s sequence, it additionally merited a closer analysis of the relationship between the lines of each stanza. In order to formulate an appropriate title summation, I was forced to examine Dickinson’s logic in placing specific lines within the same stanza, gaining a deeper understanding of how the phrases function in tandem to produce a coherent narrative.

While dissecting the similarities between lines within individual stanzas of the *same* poem proved an effective method of memorization, I quickly realized that a pure focus on similitude in my process -- which entailed two distinct poems -- had the opposite effect. My selection of the two specific poems I set out to memorize resulted from an original appreciation of their likeness. Both contain an obtrusive focus on death and feature Dickinson’s beloved robins; therefore, I surmised that the two would complement one another and serve the purpose of a dual-analysis well. While poems 153 and 182 appear alike on a surface examination, the attention to detail that memorization requires lends itself towards a far more nuanced understanding of the poems’ contrasting elements rather than those they share. Notably, in 182, the poet never refers to death by name; instead, Dickinson alludes to a state of non-existence, utilizing first person narration to position herself at the center of the text. While it is a fairly straightforward assumption Dickinson forces her readers to make in poem 182, translating “[b]eing fast asleep” to no longer living, this allusory phrasing to suggest rather than explicitly state death contrasts the pointed approach of her other poem (Dickinson 6).

In 153, Dickinson writes in the third person, positioning death as the clear subject of her phrases, and anthropomorphizing its character. As she details the qualities and actions of death, Dickinson reinforces a tone of inevitability. Her use of staccato phrasing with the series of adjectives in the penultimate stanza appears phonetically equivalent to the colloquial saying “death comes knocking at your door.” The separation between the words achieved by consecutive exclamation marks mirrors the distinct echoes of steady knocks, underscoring the formidable and unavoidable tone implied by the image of death upon one’s doorstep.

Furthermore, equivalent to the careful isolation required to master a challenging measure in a Prokofiev prelude, the sheer repetition of phrase needed to memorize Dickinson’s language is a somewhat meta manifestation of the poet’s characterization of death. The language in the final stanza of poem 153 -- which I personally titled “bird cycle” -- emphasizes the cyclical nature of life and death as Dickinson writes “[r]obin after [r]obin,” the continuum of existence endures (Dickinson 15). As I continually practiced the poem, circling back to the opening line after each complete run-through, I found that my act of recitation mirrored the untiring repetition and diligence of death. My own process of memorization took on the “[i]ndustrious” quality Dickinson assigns death in poem 153, and as a result of this effort, I achieved my primary goal:

her stanzas were stuck in my mind (Dickinson 9). The act of consistent repetition to achieve this internalization, while cyclical in nature, did not prove as monotonous. With each delivery of the poem, I noted an oscillation of tone in the second stanza. While one encounter of death in the poem emphasizes a tone of fear and formidability, a subsequent interpretation of the same stanza borders on Dickinson sympathizing with a solitary being that “[h]adn’t any playmates” and never “knew ‘his [father]’”(Dickinson 5,7). Upon this interpretation, a note of compassion rings out contrasting the otherwise sinister tone, reminiscent of the development of a sonata where the piece explores novel themes before a return to the recapitulation, where the dominant key is reinstated.

Equating Dickinson’s poems to music throughout my process of memorization provided analogous measures not only for the content but also the rhythmic flow of her stanzas. In contrast to 153, poem 182 is composed of merely two stanzas and strays from the tone of dark inevitability Dickinson crafts in 153. Considering its more compact nature and simple flow, I found poem 182 to be the equivalent of a short and sweet tune, reminiscent of a nursery rhyme. Underscoring this determination, Dickinson’s employment of a conditional clause -- stating both “[i]f I shouldn’t be alive” and “[i]f I couldn’t thank you” -- emphasizes a lighter, contemplative tone as the poet ponders her future rather than warns the arrival of death as she does in poem 153 (Dickinson 1,5). The poet’s phrasing implies that death *may* occur, depending on some circumstance; Dickinson’s use of the conditional, along with the nursery rhyme-esque flow of the poem, contribute towards a sense of lesser urgency compared to 153. Whereas Dickinson mandates in poem 153 that death will arrive “[p]unctual[ly]!” no matter her actions as a passive bystander, she instead contemplates the *possible* circumstance of her death in 182 (Dickinson 10). This stylistic contrast contributes to the overall more menacing, inevitable tone of poem 153 against the innocent, melodious nature of the second that accompanies their respective phrasing well.

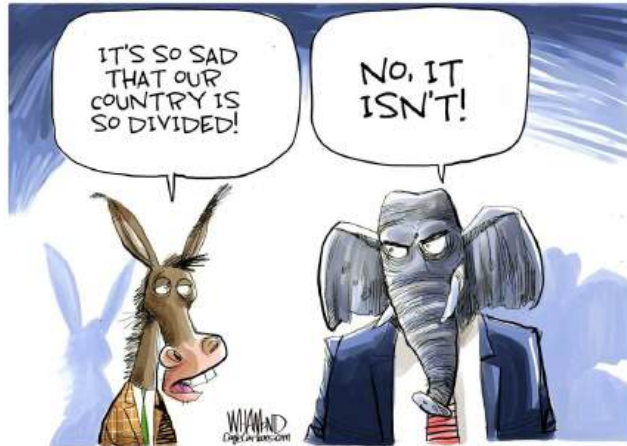
As I oscillated between memorizing poems 153 and 182, my musically-inspired technique allowed me to paint a vivid mental tableau of Dickinson’s portrayal of death. Harnessing the familiar act of musical memorization and drawing on the endless parallels that exist between music and poetry as disciplines elevated my dissection of Dickinson’s works. With music invigorating both my process and outcome of memorization, I have gained a more insightful examination of diction, complexity of tone, and analytical meaning beyond my previous engagement with the same poems. With her phrases and their semantics swimming around my subconscious mind, I earnestly report that Emily Dickinson’s poems 153 and 182 are effectively stuck in my head like tuneful ballads.

Works Cited

- Dickinson, Emily. "#153." *The Complete Poems of Emily Dickinson*, edited by Thomas H. Johnson, Back Bay Books, 1955, pp. 72.
- Dickinson, Emily. "#182." *The Complete Poems of Emily Dickinson*, edited by Thomas H. Johnson, Back Bay Books, 1955, pp. 86-87.

Liberalism Wins Most Hated Doctrine of the Year By Tessa Gumberg

Francis Fukayama's latest publication, *Liberalism and its Discontents*, defends the ideological framework of American society against the seemingly ubiquitous arguments against it.



In a society fraught with partisan schisms, a perpetual barrage of misinformation, and an overwhelmingly cynical sentiment, it has become an accepted truth that the United States is a nation divided. Despite the menacing and rapidly increasing fissure in American society, a point of harmony has been uncovered: a whopping 96% of those in a Pew Research Center study can agree upon the fact that they do not believe our “political system is working extremely or very well” (Pew Research Center) While I do not posit that this statistic feels far off from the reality of American life, it is nonetheless a disheartening number to grapple with.

Simply put, drastically opposed sides of the American political spectrum have discovered common ground in their collective detest for the very structure of our society: the doctrine of classical liberalism. While the etymological adjacents of “liberalism” have somewhat marred its true definition, the ideology at play in the skeleton of America’s social order can be defined as one that prioritizes the “foundational importance of equal individual rights, law, and freedom” (Fukayama 5). This characterization of classical liberalism, voiced by renowned political and financial scholar Francis Fukuyama, appears in the authors’ 2022 work entitled *Liberalism and its Discontents*, which tackles the salient grievances against this central order. Throughout the concise yet rich piece, Fukayama offers a stark justification for the superiority of classical liberalism as he evaluates and ultimately squashes the arguments against this structure by extremes on the right and left. As we exist in an era where unprecedented polarization is commonplace, I believe Fukayama’s publication marks a timely and effective defense of our political structure that can hopefully inspire a revived understanding and “celebration” of its foundational aspects to ameliorate this division.

Fukayama’s defense of classical liberalism builds upon his previous scholarship, notably the author’s famed 1992 *The End of History and the Last Man*, which argues for the superiority

of post-Cold War liberal democracy and denotes this doctrine as our modern world's final order. Prefacing his argument in defense of liberalism, Fukuyama states that our world is enduring a "democratic recession or even depression," and this regression exists due to dissatisfaction with the practice and application of liberal principles rather than a fault in the doctrine itself (Fukayama 21). Recent Gallup polling reveals that since 2003, America's partisan divide has increased significantly in relation to issues of the federal government "at the forefront of the political and ideological battleground" (Gallup). This statistic only emphasizes the necessity for pieces such as Fukayama's work to shed light on the most elementary principles of our nation's political structure with the hopes that understanding and assessing their strengths and weaknesses will lead to a renaissance of rationality and unity.

At the commencement of his defense of liberalism, Fukayama establishes an effective three-part argument, which he utilizes as criteria for assessing and countering the various discontents presented in opposition. His standard of reasoning traverses the realms of economic, pragmatic, and moral theory in order to provide a comprehensive dissection of the arguments against liberalism. Economically, liberal theory and property rights go hand-in-hand, and this relationship is essential to the financial growth under this order. Pragmatically speaking, Fukayama posits that a liberal order is the sole structure equipped to govern a diverse society due to its most elementary principle of tolerance. The final and most contentious aspect of his overarching rationalization concerns liberalism's moral facet: the principle of autonomous existence and choice being delegated to citizens, most explicitly evidenced by the famed line, "We hold these truths to be self-evident that all men are created equal." Undoubtedly, this fundamental concept has failed to actualize most dramatically out of the three in the doctrine's application (a dissatisfaction brought up in many of the progressive arguments against liberalism). While the crucial piece of liberal principle is choice entrusted to its citizens and dignity in the form of autonomy, the overt failures to implement this in American society overshadow any validity of the argument for its principle role. Considering this, the book warns that the issue lies within the enforcement and practice of the doctrine rather than any intrinsic structural flaws. While Fukayama dedicates a lengthy page count towards proving that equality and autonomy are central to the doctrine and-- executes it well -- it remains an unreasonable ask for those who harbor discontent against liberalism to simply accept that their grievances are misplaced without a proposed solution. It is unequivocally challenging to acknowledge that a theory is not at fault when the power structures in place fail to implement its most basic principle, lending itself to the question of how can such a supreme structure permit such a wide margin of failure.

Within the subsequent chapters, Fukayama notably delves into the history of liberalism, dissects its fundamental aspects, evaluates the arguments posed against the doctrine on both the right and the left, critiques the suggested alternatives, and concludes with a statement that the Greeks' principle of moderation may be the only way to salvage our declining democracy. Within these subsections, Fukayama grounds the ranged content by weaving in his three-part touchpoint; this effective organizational technique guides the text, helping to streamline the

somewhat esoteric language and involved content. He skillfully reiterates his thesis so that by the book's conclusion, it is undeniable that the three basic principles of liberalism -- governing a diverse society, autonomy, and property rights -- have eroded under both the far right and left control, placing classical liberalism itself under direct threat.

Supplementing his stance, Fukayama seamlessly weaves in other scholars' opinions and words throughout the text, augmenting his existing credibility as an author and theorist. By including a breadth of references to other politicians and academics, Fukayama diversifies the tone and flow of his prose, crafting a resounding voice that compliments the technical subject matter.

In my opinion, not only does Fukayama present a conscientiously structured defense, but he crafts an equally effective damnation of the alternatives. As Fukayama explores the proposed substitutions for classical liberalism by the left and the right, he evaluates these alternatives to the criteria established by his three-part original argument. This consistent thread of economic, pragmatic, and moral standards throughout the text truly drives the point home that there are, in fact, no genuine alternatives to classical liberalism.

Liberalism and its Discontents can be simplified to a treatise as to why the critiques of our current order are debased and unfit for consideration. Fukayama unearths the fundamental truths that compose the structure we live under, and while his work does not inspire a nouveau optimistic outlook on our current system, it certainly evokes a desire to dissect the ideological anatomy at play. I couldn't agree more with the premise of his book: it is crucial to understand the principles of our society if we hope to not only survive but excel under this order that, as Fukuyama concludes, is the "end of mankind's ideological evolution" (Fukayama 98). Liberalism is not going anywhere, and this work takes significant strides toward convincing the discontented that they should not wish it to, either.

Works Cited

Fukuyama, Francis. Liberalism and Its Discontents. Straus And Giroux, 17 Aug. 2024.

Gallup. "Update: Partisan Gaps Expand Most on Government Power, Climate." Gallup.com, 7

August 2023,

[news.gallup.com/poll/509129/update-partisan-gaps-expand-government-power-climate.as](https://news.gallup.com/poll/509129/update-partisan-gaps-expand-government-power-climate.aspx)

[px.](https://news.gallup.com/poll/509129/update-partisan-gaps-expand-government-power-climate.aspx)

Pew Research Center. "Americans' Dismal Views of the Nation's Politics." Pew Research Center

- U.S. Politics & Policy, Pew Research Center, 19 September 2023,

[www.pewresearch.org/politics/2023/09/19/americans-dismal-views-of-the-nations-politics/.](https://www.pewresearch.org/politics/2023/09/19/americans-dismal-views-of-the-nations-politics/)

On Morality: Is Morality Purely a Construct of Cultural Tradition?

By Daniel Yu

Preface to the Reader.

The task of writing philosophy is two-fold. The first hurdle is discovering one's thoughts, unthinkingly wading through content without view of a clear path ahead. The second is the hurdle of expressing these thoughts in an alluring yet precise manner, slowly yet surely leading the reader into the mind of the philosopher. Regarding the latter, Descartes comes to mind as a particularly seductive writer, his First Meditations opening the world of philosophy to me. One of the aspects I admire most about Descartes is his originality of thought, and so in light of this I am proposing a unique ground-up argument, rather than working off the framework built by other moral philosophers. However, in replying to objections with stock responses, I shall utilise well-accepted solutions proposed by others (insofar as I agree with them), as there is no use in entirely ignoring philosophical progress made in this field.

When dealing with a question such as this one that dips into both analytic and continental philosophy, holding dear openness and approachability will prove fruitful and enjoyable. Despite this, do note that language used in this essay will have specific philosophical connotations rather than colloquial ones in mind. However, any particularly niche terms will be pointed out and defined.

I. Introduction

Morality is a word often used yet often misunderstood in today's society. Angered claims of 'such and such' being immoral are rife within the armchair politics that so plagues the centuries of study that philosophers have laboured over. Hence, we must pinpoint the question to avoid making the same mistake.

The word morality derives from the Latin "moralis" meaning 'pertaining to customs'. The word tragedy derives from the Greek "tragos" meaning 'goat' and "ode" meaning song. For now you may think that the former of these roots makes more sense than the latter, however by the end of this essay I hope that I will be able to convince you of the exact opposite.

The heart of this question comes not in the latter half that reads "construct of cultural tradition" but is instead found in the single word "morality". The question is, in fact, asking us, "What is the source of morality?" as answering this will subsequently provide an answer to the original question. Broadly, the essay will flow as follows:

1. Defining morality and a discussion of its possible sources
2. The issues of a morality contingent on cultural tradition or other contingent concepts
3. Founding morality in a deity

This essay will reflect that morality without God is unfounded, misguided and cruel, concluding that morality is entirely separate from cultural tradition. The core of this essay is the

claim that morality exists only if it is objective. As a consequence, I will also show that the initial appearance of morality being a construct of cultural tradition is a misinterpretation of the nature of morality, confusing it with a synthesis of emotion and opinion.

II. A discussion on the nature of morality and its possible sources

There are two facets to my definition of morality:

- (i) Morality is defined by its normative use that determines what people ought and ought not to do, and should be followed by all rational people
- (ii) If morality exists, there also exists an objective moral standard

Establishing such carries much of my argument on its shoulders, so defending both claims seems required. The first claim is relatively uncontroversial - it is almost intuition to say that if we deem an action immoral, we ought not to do it, and if an action is moral or morally neutral then it is at least morally permissible.

The only common objection that may arise is that if something is moral, that does not mean we ought to do it. This issue is tackled by introducing the term 'morally permissible', as surely one cannot deny that an immoral action is something that one ought not to do. This is a contradiction by how we have defined the terms 'moral' and 'immoral', and more importantly is appalling to our intuition. Thus, I believe we can accept claim 1.

The second claim is more controversial yet appears commonsensical at first glance. For something to exist, this often means moving beyond semantic and epistemological realms into the metaphysical realm. To clarify this, I will provide an example.

If I were to ask you whether a word exists, after a short think, I believe you would provide the following response:

"The word itself may not exist, but the thing or descriptive it refers to certainly exists."

This intuition makes sense, as if words exist then this means that humans are able to bring non-physical things into and out of existence when we make or discontinue languages. And so, it appears to be in our human nature to instinctively believe that for something to exist, it must exist beyond its semantic use, in this case, the words "slowly" and "good".

Existence going beyond epistemology is more difficult to verbalise - since epistemology refers to our theory of knowledge, perhaps consider whether the knowledge of our earlier words of "slowly" and "good" exists. Again, I think you may conclude that knowledge is too intangible to exist - as if they were to exist, then they would appear and disappear from existence without us noticing (for we would no longer have the knowledge), with us only recognising our lack of said knowledge. Perhaps the knowledge exists as particles flowing in our brain, and yet we find ourselves much more tentative in asserting the existence of knowledge outside of this physical

state. However, it is not a problem to state that knowledge exists in this physical form, but this is no longer within the realm of epistemology.

Therefore, we consider existence in the metaphysical realm to be the sufficient condition for something to exist. This is expected, as metaphysics is the study of reality and thus concerns existence itself. For how could I deny that the chair I am sitting on is part of reality and thus exists, just as much as the desk that my elbows rest upon? Now I must show that metaphysical existence entails an objective moral standard.

To do this, consider what makes up 'morality'.^[1] I think that a reasonable definition would be that it is a set of moral evaluations (i.e. deciding whether something is moral or immoral) covering every possible scenario in every possible world^[2], for there cannot be a situation that lacks any moral evaluation whatsoever. For example, an evaluation of whether murdering someone in a yellow subway station on Tuesday is either moral, immoral or morally permissible. Since these exist metaphysically, there must at all times exist a set of these evaluations that exist objectively^[3] outside of human influence, as claiming that moral standards are not non-physical is undoubtedly a difficult and unpopular stance to take.

Here, the title of the paragraph may be misleading, as there are not possible sources (plural) of morality, but merely one, being a deity (a possibility that I aim to prove). Since it is clear how a deity would be a possible source of morality, I believe it is more interesting to consider how all non-theistic sources of morality collapse into one bundle of subjectivity before tackling this in my next paragraph. Now, considering the original question, possible suggestions of sources of authority include: it is a construct of cultural tradition, emotion^[4], and upbringing. So, let us consider how we come to develop (and therefore believe) moral beliefs into a 'morality'.

If I see a cat being killed on the street, I am immediately filled with sadness and anger, which translates to my moral belief that killing cats is wrong. In my attempt to rationalise this, I return to first principles, such as "killing without reason is wrong". Here, it is clear that emotion takes the reins.

Then let us consider upbringing: when we do something wrong, we see our carers get angry or upset at us. Then in response, we subconsciously attempt to rationalise their emotions as we assume that these carers feel these emotions for a reason. From here, either we look back upon our first principles or form new first principles, such as "taking another child's pen is wrong". Hence, we see the process of emotion, rationalisation, and checking back on first principles yet again. Cultural tradition is the same.

By the term cultural tradition, I will take it *prima facie* to mean practices of a group that have been sustained for an extended period, thus resulting in something that resembles a standardised set of beliefs or values. Consider how these practices were first established as the norm. I believe it is fundamental human nature to seek what makes us happy and that makes others happy, and even if this were not the case, it seems almost common sense that societies act in a way that they believe to promote the prospering of that society or the individuals within it.

Imagine a society within which the first murder has just been committed; the immediate response shall not lie in rationalisation but in an emotional response of shock and anger. From here, rationalisation from the emotion occurs, thus forming the societal norm of murder being something that ought to be avoided.

I look upon these several cases and note a pattern - the initial reactions that form or encourage us to consider first principles are always emotion, which is certainly subjective. Thus, hopefully I have been able to convince you that:^[5]

- 1) The formation of moral beliefs comes from either:
 - i) A non-theistic source, including cultural tradition, and is dependent on the collective emotional response of society
 - ii) People's decision making that is composed of their own response to a subjective moral outline
- 2) As emotion is subjective, if a non-theistic morality were to exist, it would be subjective and contingent on either:
 - i) emotional response or our intellectual parsing of our emotional responses

III. The issues of a morality contingent on cultural tradition or other contingent concepts

Establishing the above leaves me with the task of convincing you of two things:

1. A 'subjective morality' is the equivalent of no morality at all
2. A theistic source of morality can provide an objective morality

This section shall tackle the former, with my final section tackling the latter. To do so, I will provide a two-pronged attack on subjective morality, highlighting its distinction to objective morality. The first shall be the implications if we accept that morality is subjective, and the second is a more direct strike to its core by showing that 'subjective morality' holds no substance as a concept.

Suppose I was to ask you whether killing an innocent baby is wrong. I assume you would reply emphatically that it is! Next, I ask you whether killing someone's pet is wrong. Again, the reply would swiftly be an affirmative "yes!". And yet, what does accepting subjective morality entail? How could you say that killing an innocent baby is wrong if everybody's idea of 'wrong' is different to everybody else's? Surely instead of saying that it is wrong, your statement of disapproval becomes something like saying "Boo^[6], murder"^[7]. Here, we can no longer call it a moral evaluation as it becomes a personal opinion.

If something being moral or immoral is not fixed and varies between people, I ask why punishment is just. You may think that whatever the majority of a society believes to be moral should be viewed as moral, and vice versa for immoral actions that we are then justified in punishing. Then I ask you, was slavery moral? Was the Holocaust moral? I suspect you find

yourselves itching to reply with a profound "No!"; yet, with subjective morality, doing so would simply be stating an opinion.

Perhaps the strongest case against subjective morality is that subjective morality cannot 'exist'^[8]. Here, it would be helpful to remind yourself of the definition of morality established earlier, particularly the form it takes as moral evaluations. Let us consider an example of a moral evaluation for subjective morality.

I ask you if stealing from a homeless man is moral. As a moral subjectivist, you reply, "In my opinion, it is immoral". Yet, this cannot be a moral evaluation. This is because many events we must morally evaluate involve ourselves, someone else, or others entirely separate from ourselves. If our moral belief is subjective, it is wholly contained within ourselves, so we cannot apply it to situations involving others with their own subjective moral beliefs. For example, what if the homeless man believed stealing from him was moral? Here, subjective morality is placed in limbo, where doing the action is both moral and immoral.

Therefore, subjective morality is lessened to different sets of opinions. I believe that labelling peoples' sets of opinions as individual moralities is absurd and unnecessary, for we have reduced morality to something almost nonsensical, as if it is a balloon lacking tethering to a metaphysical ground. When we consider a subject such as morality that holds the lives of countless people in its hands, we cannot settle for anything less than solid earth that we can reliably feel beneath our feet. Yet what is subjective morality but a sea of ever-changing, inconsistent views?

IV. Founding morality in a deity

This section shall be brief, yet it is necessary to show that theistic approaches may provide us with a source of authority. Note that the task here is not to prove that a theistic source is the only possible source of morality, as this has been done previously by showing that non-theistic sources cannot be possible sources of morality. Here, the task is simply to show that it is logically possible for morality to come from a theistic source, as this allows us to rule out non-theistic explanations for what morality is whilst accepting that morality may still exist. Let us consider the properties of the God of classical theism. This God has the following properties:

- (i) All-knowing
- (ii) All-powerful
- (iii) Completely good

This God would be able to create moral standards as God is completely good, knows what such moral standards would be and is all-powerful, allowing Him to bring these standards into metaphysical existence. Therefore, it is logically possible for morality to be a 'construct' of a deity. The truth-formal logic appears as follows:

P1: Morality must be objective if it were to exist

- P2: Non-theistic explanations of morality cannot provide an objective morality
P3: Theistic explanations of morality can provide an objective morality
P4: The sets of Theistic and Non-theistic are exhaustive for possible explanations of morality^[9]
C1: A correct explanation of morality cannot be found in non-theistic explanations if it were to exist
C2: A correct explanation of morality must be found in theistic explanations if it were to exist

Finally, we see the result of C2 that this section has aimed for. Notice that it is not an issue if other deities do not share similar properties, as the aim here is to show that it is still logically possible for morality to exist if we rule out all non-theistic sources.

V. Conclusion

We have reached a conclusion that has allowed us to progress, yet still feels disappointing. I have provided a trail of thoughts that has led us to believe that if morality were to exist, it must be founded within a deity. In turn, we have answered the topic question, showing rather conclusively that even linking any source of morality to cultural tradition is unfounded adherence to emotion and so is mistaken.

The best way to view cultural tradition is to observe it as a telescope. When we peer through the telescope up at the night sky, we see a beautiful sculpture on the other end, desperately trying to recreate it ourselves with earthly paints and paintbrushes. What we obtain is our painting, and still this telescope remains entirely separate from the night sky it points towards. Likewise, we use the telescope of cultural tradition to helplessly catch a glimpse of the objective morality that lies so far away, and we illustrate what we see into an imitation that satisfies us. Cultural tradition can never hope to contribute to what morality is, but it can at least help in sketching its rough outlines.

A Defence of ‘On Morality: Is morality purely a construct of cultural tradition?’

I An Outline of the Raised Objections

In both reviewing my original paper ‘On Morality’ and receiving objections against several parts of the argument, I have discovered flaws in my argument and/or wording that ought to be fixed. These flaws have left parts of the argument open for valid objection due to lack of clarity, though I shall also tackle the objections against which my argument resists. I shall start by expressing these rebuttals as strongly as I can. The triad of theistic attributes is inconsistent, both internally and externally:

- The problem of evil causes us first to question the consistency of the triad, and second to question any morality that arises from such a god. For without the guarantee of omnibenevolence, why should we regard a set of ‘Commandments’ (i.e. a morality set out by a god) as our given moral system?

- The omnipotence paradox causes us to question further the consistency of the triad.
- Euthyphro's dilemma causes us to doubt a god as the true source of morality

The groundwork for various objections has now been created. First, the possible origin of morality proved in my original paper arose from a God that had been given the triad of aforementioned attributes, and hence proving the impossibility of said attributes may question the validity of this (let us name the case from which the version of morality as defined in my original paper may be created the 'possible case'). Second, Euthyphro's dilemma (if it caused us to conclude that there is an objective morality separate to God) would force us to find a different origin for said morality.

I address these by isolating considerations on the validity of my definition of morality (analysing the possible case), and then offering a defence of the triad of theistic attributes.

II. On The Purpose of the Argument

By providing God with the triad of theistic attributes (i.e. addressing the God of Classical Theism) it may at first appear that this opens the paper up to a wealth of objections, namely all those proposed against following properties:

1. Omniscience
2. Omnipotence
3. Omnibenevolence

However this is not the case. Let us display this by considering the intention of the argument carefully. The essay broadly takes form through two halves culminating to respond to the given question of "Is morality purely a construct of cultural tradition?" with a profound response of "No!". The two halves can be defined as such:

1. Providing an intuitively appealing definition of morality (let us name a morality defined in such a way G-morality) that illustrates the impossibility of cultural tradition fulfilling said definition.
2. Demonstrating that such a morality is not absurd, in that there is a given set of circumstances through which they can arise, which may be expressed as thus:
 - A god with the triad of properties (i.e. a god defined with the attributes used in the paper) is not impossible
 - From such a god, it is possible for the definition of morality provided to be fulfilled

We must analyse this carefully to see how it deals with objections against the triad, and in fact allows us to ignore almost all of these objections entirely. The crossing-of-the-ocean comes in observing that the objective of half 2 is not to say that there necessarily exists a G-morality, rather it is to display the validity of the definition that had been asserted earlier on in the paper.

Consider an example argument to the title question that takes the following form:

1. Provide an intuitively appealing definition of morality
2. Display that cultural tradition fails to fulfil the requirements for such a definition
3. Conclude that as a result, morality cannot arise from cultural tradition

The weaknesses of this argument illustrate why the argument found within my paper stands. The crucial objection is that this argument has failed to offer a defence of the definition provided, thus failing to account for the possibility that cultural tradition's failure to meet the criteria is not a failure of cultural tradition, but a failure of the criteria itself.

Suppose I define a 'good' square as one that has 3 sides, and you ask whether a given square you have drawn is good. Without ever examining your square, I can conclude that it is 'not-good'. The square may well be good under different circumstances, and yet with the definition provided it is not. In this case, we can determine that the given definition should not be a 'valid' one as there is no possible entity that could satisfy the conditions, and hence a conclusion of a square being 'not-good' under the given definition is not meaningful.

Returning to the faulty example argument, one central reason as to its failure is that it fails to display a case for which the definition could feasibly be achieved, hence failing to provide reason to believe that the definition is valid at all.

As a result, we discover the key advantage of my argument that allows us to avoid most arguments against the triad. Unless it is the case that it is logically impossible for the set of circumstances that I provided to occur, then there is a possible entity for which G-morality is reached, allowing us to conclude satisfactorily that the definition provided is valid (as long as it remains intuitively appealing), and thus illustrating that the failure of cultural tradition does not stem from a failure of the definition. For if we had provided a valid definition for a 'good' square, it is meaningful to say that a square is 'not-good'. Equivalently, it is meaningful to say that cultural tradition fails to meet the criteria for G-morality. Let us review what we have established:

- It is not necessary to show that G-morality exists, rather that there is a possible case in which it is achieved
- The purpose of my argument is to display that such a case can arise, ignoring how improbable it may be (as long as it is probable enough to remain intuitively appealing)

- Through the combination of the above, we conclude meaningfully that morality can never be a construct of cultural tradition

III. Logical Problems against the Possible Case

From here we are left with one task: showing that it is not logically impossible for the properties I have attributed to exist simultaneously. Again note that it is only the logical formulations of objections that may damage the argument, for these are the only cases that may eliminate the possible case of G-morality required for our definition to be valid. These objections come in two forms, either internal logical cohesion (name this internal cohesion) or logical cohesion between different attributes (name this external cohesion). I now proceed by considering the logical objection to each case individually.

3.1 Omnibenevolence

Objections against omnibenevolence tend to be on external-cohesion, particularly on the Problem of Evil (combining omnibenevolence, omnipotence, and omniscience). Alvin Plantinga has offered a well-accepted solution to the logical problem of evil, known as the free-will defence. Let us first borrow Plantinga's definition of transworld depravity outlined in "God, freedom and evil":

"A person P suffers from transworld depravity if and only if the following holds: for every world W such that P is significantly free in W and P does only what is right in W, there is an action A and a maximal world segment S' such that.. (Plantinga 48).

- *S' includes A's being morally significant for P*
- *S' includes P's being free with respect to A*
- *S' is included in W and includes neither P's performing A nor P's refraining from performing A*
- *If S' were actual, P would go wrong with respect to A "*

In much less formal language, the core concept is that an agent suffering from transworld depravity will perform at least one morally wrong action in any possible world he is created. I shall formulate his argument which utilises transworld depravity as follows:

- P1: There are possible worlds that even an omnipotent, omniscient, and omnibenevolent being cannot actualise
- P2: It is possible that morally free agents have significant non-compatibilist moral free will (SNCMF)
- P3: It is possible that humans are morally free agents • Conj: It is possible that humans have SNCMF
- P4: Non-compatibilist moral free will is significant iff every moral agent is given the choice to do either good or bad activities

- P5: It is possible that all morally free agents suffer from transworld depravity • Conj: A world with SNCMF and without bad activities is not possible
- P6: It is possible that a world with SNCMF and both good and bad activities is more optimal than a world without SNCMF and only good activities
- P7: An omniscient, omnibenevolent, omnipotent (name this entity D) being always realises the optimal world possible
- Conj: It is possible that D realised a world with SNCMF and both good and bad activities
- Conclusion: There is no logical inconsistency when an omnipotent, omniscient, and omnibenevolent entity creates a world of free creatures who do evil (SNCMF + bad activities)

This argument successfully deals with the logical problem of evil (with the weight of rejecting this argument being accepting the compatibilist notion of free will, an acceptable stance though the debate then hinges around free will rather than God's omnibenevolence). Thus, we have dealt with the first objection to the consistency of the triad.

3.2 Omnipotence and Euthyphro's Dilemma

Both of these objections can be dealt with swiftly as under examination they collapse into false dilemmas, hence I shall address them in the same section.

Omnipotence paradoxes tend to reduce to a question of the definition of omnipotence, and come in the form of internal-cohesion objections. The two most common definitions are as follows:

1. The quality of being capable to realise any outcome that is not logically contradictory (i.e. anything with a referent)
2. The quality of being capable to realise any outcome, including that which is logically contradictory

Accepting definition 1 simply ends the paradox. Although *prima facie* it may appear that this is limiting a god's power, I would urge you to question whether preventing a god from 'creating a square circle' or 'creating a rock so heavy that (h)e could not lift it' either affects the argument, or limits a god's power at all (my intuition does not think so, for the only world in which I can envision a 'square circle' being possible would be one in which both of these words are redefined, removing the logical impossibility, and any other world being outside of my understanding).

Euthyphro's dilemma is commonly posed in the following way: There are good and just things. Either (addressing the God of classical theism):

1. These things are good and just because God commands them
2. God commands these things because they are good and just

For my argument to stand, one must either accept option 1 or dismiss the paradox altogether. I shall choose the latter, instead opting for a third choice: God does not invent nor conform to moral order, rather His omnibenevolent, unchanging nature is the standard for moral value. This successfully deals with Euthyphro's dilemma without giving up any ground or damaging my argument, as neither is morality separate from God in this case, nor does it become wholly dictated by God. Hence the argument still stands.

IV. Conclusion

With both wording corrections and responses to given objections having been made, my original argument still stands. However, it ought to be noted that this argument shall still be appealing only to those that accept the possibility of some given conditions:

1. The definition of morality provided is intuitively appealing
2. It is plausible enough that the god defined in the paper exists so that we do not need to remove the intuitive appeal of the definition of morality in hindsight
3. Non-compatibilist moral free will is an acceptable (/not unreasonable) stance to take
4. The parts of the original argument that have not been objected to are accepted

With these given conditions, the conclusion that morality is not purely a construct of cultural tradition still holds true.

Footnotes

^[1] The term is used very loosely here for simplicity

^[2] To consider possibility and necessity, I am using the possible worlds approach

^[3] The terms objective and subjective will be defined like this: An objective thing is one that if humans were to stop existing, they would still exist. A subjective thing may be changed or influenced freely between different individuals.

^[4] A. J. Ayer's writings on ethical emotivism are worth reading on this

^[5] Thanks to those that confirmed these intuitions, namely those in Westminster School

^[6] I have been inspired by Alex O'Connor to use such a term (boo meant in its disapproving connotations, not those of a ghost)

^[7] A.J. Ayer's emotivism follows a similar path

^[8] 'Exist' used in how it was defined earlier

^[9] Law of excluded middle shows that this is true, and both are negations of each other

Works Cited

Plantinga, Alvin. God, freedom, and evil. Eerdmans Publishing Company, 1977. Accessed 17 June

Fine-Tuning Neural Networks for Enhanced Emotion Detection: Insights from CNN, ResNet-50, DenseNet-121, and ViT Models By Amogh Singh

Abstract

This paper explores how to effectively choose and fine-tune different Neural Network models to detect emotions. Specifically, it highlights how to use the Keras Sequential API CNN, ResNet-50, DenseNet-121 and ViT models, and fine-tune them so that they yield the greatest accuracy. Emotion detection has several applications, including psychology research, public instruction, and security. This paper uses a publicly accessible dataset from Kaggle to train and assess these models in our studies. After a thorough investigation, the paper proposes recommendations to improve the accuracy of the chosen Neural Network Model.

Introduction

Emotion detection using Artificial Intelligence (AI) is a growing field of study that has many potential uses, such as psychology studies, public education, and safety (Gajarla and Gupta). An example of this is where an individual is driving a car and the model detects that the driver is angry or upset. Then, the car can send the driver an audio message to drive carefully, play a happy song, or turn on a massaging mechanism in the seat. This can potentially save many lives since motor vehicle accidents are one of the leading causes of preventable death (Girasek 455-465; Koh et al. 501-506; Ray et al. 676-680). This can also be integrated into the study of psychology, to find the reactions of individuals due to a certain stimulus, with less bias than if an additional person was present. As a result, scholars will find it much more feasible to conduct psychology studies and they will be able to yield data that will be less suspect to bias (Pronin, Lin, and Ross 369-381; Pronin, Gilovich, and Ross 781-799; Rosenthal and Fode 491-511)

This paper conducts emotion detection through image classification, a critical component in computer vision that trains algorithms to recognize and categorize objects in digital images. Image classifiers learn from labeled training data, where each pair of images and the associated class label are grouped together. When researching a complex problem like emotion detection, a fairly large dataset is necessary to yield accuracy, due to the different nuances that incorporate different emotions. Preprocessing is an essential part of handling the dataset. For example, problems occur when the angle and exposure change a picture, necessitating the need for powerful classifiers to adjust. Preprocessing, however, somewhat fixes the challenges of dimensionality, which causes pixels to contribute to a large feature space.

Furthermore, choosing a suitable model is very important when it comes to image classification and certain aspects carry great importance when choosing the most optimal one. Some considerations are accuracy, use of computational resources, the ability of generalization to avoid underfitting or overfitting, and scalability for the dataset. In addition, it is crucial to fine-tune the model and make sure the model's architecture aligns with the dataset's characteristics, which include image resolution and diversity. Some ways to fine-tune a model could be to change the number of epochs, batch size, image size, patch size, validation split, etc.

This will further enhance the performance of the model and ultimately help achieve the goal of detecting emotion detection through AI (Qayyum et al). This paper compares the results of different models and adjustments to their architecture to detect emotions through image classification.

In addition, most CNN models will have convolutional blocks. These consist of one or more convolutional layers that extract features from the input image. The convolutional layers are typically followed by one or more pooling layers, which reduce the spatial dimensions of the feature maps while retaining the most important information (Dai).

Literature Review:

Neural Network Models:

This paper will compare the following models: Keras Sequential API CNN, ResNet50, DenseNet121, and ViT.

Convolutional neural networks (CNNs) are specialized deep learning structures that are primarily designed for handling grid-like data, especially images. To do this, CNN starts by breaking down the images into smaller parts. The convolutional layer uses learnable filters, like sliding windows, to scan the image. These filters start by detecting simple features and later scan for more complex patterns. With each convolutional step, flexibility is added by applying a non-linear function, which is most often a Rectified Linear Unit (ReLU). This step allows the network to store similarities between features. To decrease the computational burden, pooling layers are implemented. These layers reduce the size of the image while maintaining crucial characteristics (Koonce and Koonce 63-72; O'Shea and Nash).

Another neural network model used is ResNet50, which was created by Microsoft Research in 2015. It is a variation of the mainstream ResNet architecture with 50 layers which allow it to learn structures deeper than before while avoiding vanishing gradients. ResNet50's architecture is separated into four major components: convolutional layers, the identity block, the convolutional block, and the fully linked layers. The convolutional layers extract features from the input image, which are then processed and transformed by the identity and convolutional blocks before being classified by the fully connected layers. ResNet50 was trained on the enormous ImageNet dataset and achieved an error rate comparable to human performance, making it an effective model for a variety of image classification tasks including object detection, facial recognition, and medical picture analysis. It has also been used to extract features for a variety of other applications, including object detection and semantic segmentation. (Li and Lima).

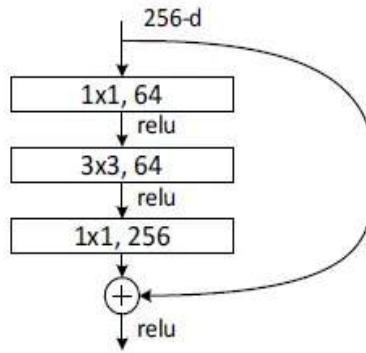


Figure 1. Residual Learning: a building block

DenseNet-121 is a convolutional neural network with dense connectivity across layers that improves information flow and feature reuse. It consists of 121 layers organized into four thick blocks, with transition layers between them. Each dense block contains several convolutional layers, each of which receives input from all preceding layers in the block via feature concatenation. This dense connection architecture improves feature transmission while addressing the vanishing gradient problem in deep networks. DenseNet-121 starts with a 5x5 convolution and maximum pooling layer, then dense blocks of 6, 12, 24, and 16 layers, respectively. Convolution and pooling are used as transition layers between dense blocks to reduce spatial dimensions. The network finishes with global average pooling and a softmax classification layer (Zhu and Newsam; Li 263-276).

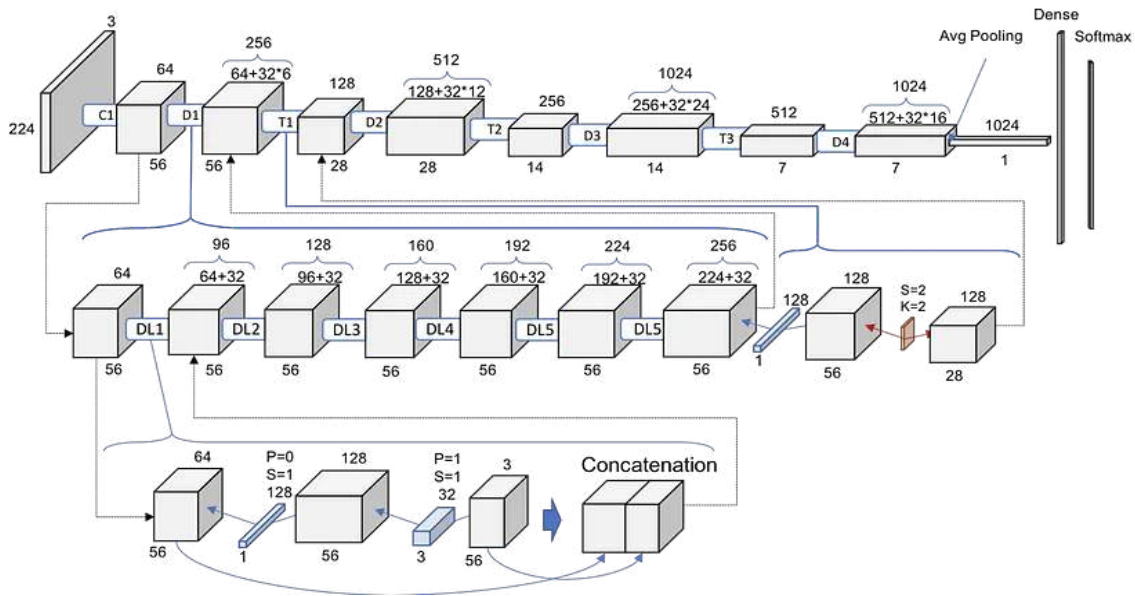


Figure 2. Full schematic representation of ResNet-121

ViT is a neural network model that is used for image classification. It splits an image into small patches and places them into a sequence. The model consists of multiple transformer layers comprising multi-head self-attention and feed-forward networks. The network starts with a linear projection of these patches and follows with position embeddings. The transformer layers handle such embeddings with self-attention mechanisms, which allow for effective feature extraction and aggregation. The model finally returns a classification token and a softmax layer for final predictions. (Han, Kai, et al; Xu et al. 33-62).

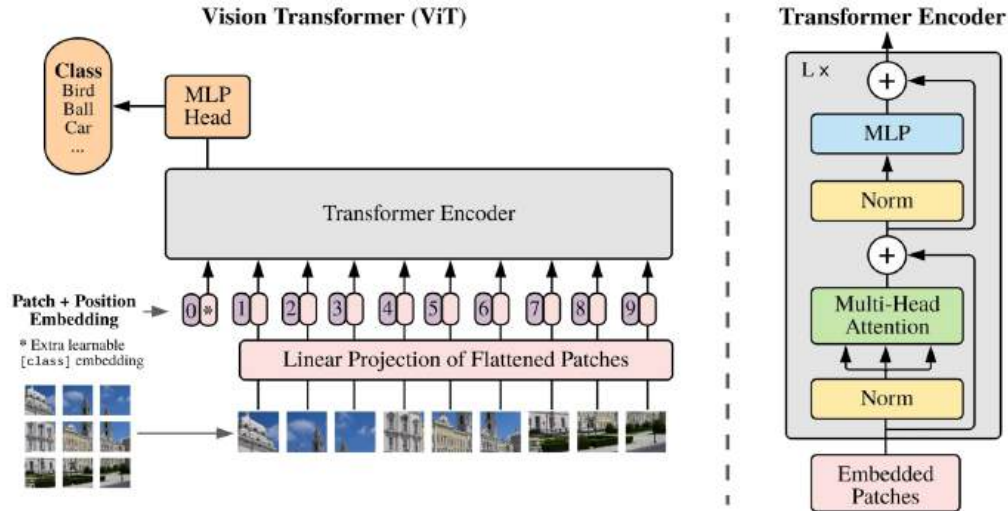


Figure 3. Overview of Visual Transformer model

Methods:

Data Collection:

This paper will be using the FER2013 Dataset, which consists of 7 classes:

Class Number	Class Name
0	Angry
1	Disgust
2	Fear
3	Happy
4	Sad
5	Surprise
6	Neutral

The training set has 28,709 images and the test set has 3,589 images, with each image being 48x48 grayscale. (Sambare)

Overview of Plan:

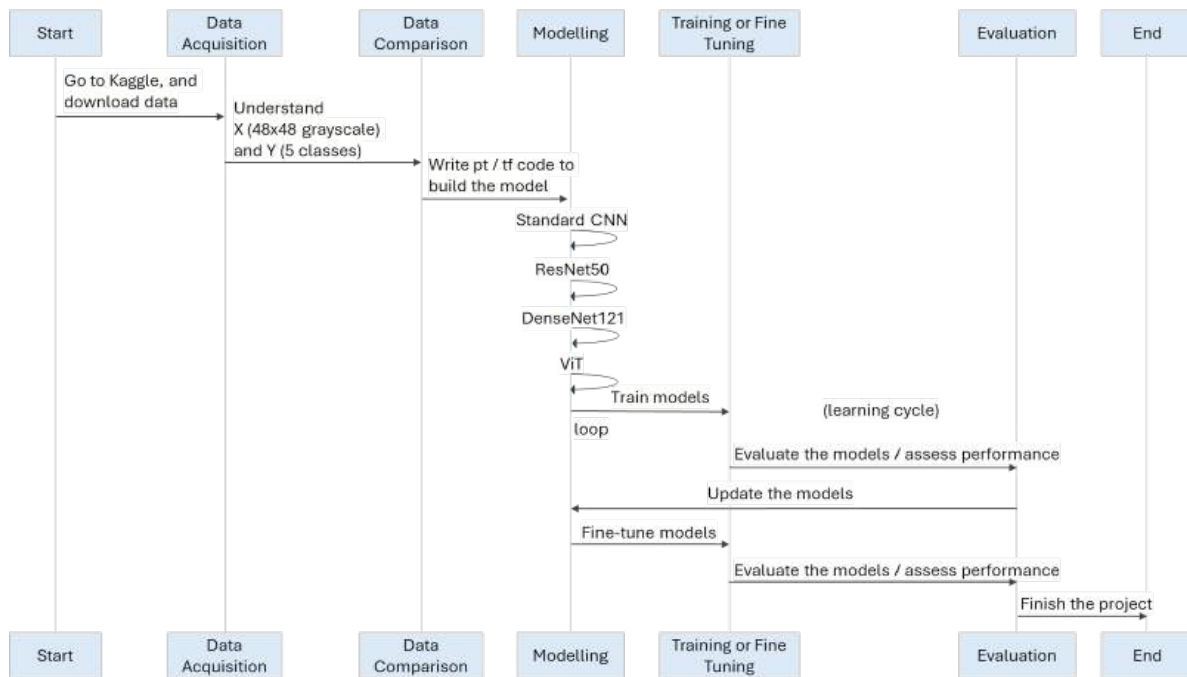


Figure 4. Project plan for experiments

Data preprocessing:

This project starts by downloading the dataset and ensuring the images are correctly placed in their corresponding folders. Then, the following folders are combined as their corresponding images are extremely similar: fear and surprise, anger and disgust. Once that is done, the pickle module is imported to open the training and test data file separately in read-byte mode. Then, the training and test data is loaded from their corresponding files. Afterward, the training data is separated into features (x_{train}) and labels (y_{train}), resulting in the test data features (x_{test}) and labels (y_{test}). Subsequently, the first channel of the training and test images is kept so that they can be converted to grayscale. Following this, LabelEncoder is imported for encoding labels, which packs the processed training and test data into a tuple. Finally, the project opens a file in write-byte mode to save processed data and saves the processed data to the file using the highest protocol available.

Then, the processed data file opens using read-byte mode, and the processed data loads from the file. Afterward, the model unpacks the loaded data into training and test sets with their labels and normalizes their data by scaling pixel values to the range 0 to 1. Tensorflow is

imported to convert these labels to a one-hot encoded format. After this, `train_test_split` from `sklearn` is imported to split the training data into a smaller training set and validation set. The validation set is set 10% of the original training data (Albahra et al; Jaramillo-Morán et al.)

Models:

Keras Sequential API CNN

The Keras Sequential API CNN starts by first flattening the input images from 48x48x1 to a 1D array. Following this, the model consists of 3 dense layers with units of 256, 128, and 64, all with ReLU activation, each followed by a dropout layer to prevent overfitting. Then, a separate dense layer is placed, with a total of 6 units along with a segment for softmax activation. Following that, model is compiled using an Adam optimizer, categorical crossentropy loss, and accuracy metric.

ResNet50 model

This model uses three functions to build this model: `identity_block`, `conv_block`, and `build_resnet50`. The identity block is the default block employed in ResNets and represents the scenario where the input activation and the output activation have identical dimensions. The `identity_block` function works to create an identity shortcut, `conv_block` implements a convolutional shortcut, and `build_resnet50` constructs the ResNet-50 model by adding all of the convolutional and pooling layers, as well as the residual blocks for prediction.

The `identity_block` takes the input tensor, kernel size, and filters as parameters and starts by unpacking the filter sizes. This function has 3 convolutional blocks. The first one reduces the dimensions by integrating 1x1 convolutions and the second one contains the main convolution layer with the specified kernel size. The first two blocks apply batch normalization and ReLU activation. The third convolutional block has a 1x1 convolution to restore the dimensions and then applies batch normalization. After the convolutional blocks, the function adds the input tensor to the output tensor and applies the ReLU activation to the summed output.

The `conv_block` takes the input tensor, kernel size, filters, and strides as parameters and starts by unpacking the filter sizes. After, it applies a 1x1 convolution with strides followed by batch normalization and ReLU activation. Next, the main convolution layer is defined with its kernel size, followed by batch normalization and ReLU activation. After that, it applies a 1x1 convolution for restoring dimensions, followed by batch normalization. The function also contains a shortcut path with downsampling, along with batch normalization. The shortcut is added to the output tensor and then ReLU activation is applied.

The `build_resnet50` takes the input shape and number of classes as parameters and starts by defining the input tensor. After that, add zero padding and 5x5 convolution with stride 2, followed by batch normalization and ReLU activation. Zero padding is again added, along with max pooling with stride 2 for downsampling. There will be 3 Residual blocks, where each will have a convolutional block with the specified strides and filters, along with their corresponding

number of identity blocks. Finally, the function applies global average pooling to limit each map to a singular value and adds a fully connected layer with softmax activation. Finally, instantiate the model. The model compiles the model using an Adam optimizer, categorical crossentropy loss, and accuracy metric.

DenseNet121

The model consists of three functions: `dense_block`, `transition_layer`, and `build_densenet121`. The `dense_block` function makes sure each layer's output is joined with the input. The `transition_layer` reduces the limits of the number of features and maps and downsamples input. The `build_densenet121` constructs the densenet 121 model with its convolutional and transition layers and dense blocks.

The `dense_block` function takes the number of layers and growth rate as parameters and then loops the concatenation process for each layer present. This concatenation process starts with batch normalization and ReLU activation. Following that, 1x1 convolution is added to increase the feature maps. Then, the output goes through batch normalization and ReLU activation, after which a 3x3 convolution is applied to extract the features. Finally, the input tensor will be concatenated with the output tensor and once the loop is completed, the output is returned.

The `transition_layer` takes the compression factor as its parameter and starts with computing the number of filters for the transition layer. After that, the output goes through batch normalization and ReLU activation. Then, a 1x1 convolution is added to reduce the number of filters and pooling is averaged to downsample the feature maps.

The `build_densenet121` takes the input shape, number of classes, growth rate, and compression factor as parameters. The function starts with defining the input tensor and follows with the initial convolutional layer. The convolutional layer contains a 6x6 convolution with stride 2, followed by batch normalization and ReLU activation. Then, max pooling is conducted with stride 2 for downsampling. Next, 2 dense blocks are created, one with 6 layers and the other with 12 layers, both of which have a transition layer to reduce dimensions. After that, apply global average pooling and add a fully connected layer with softmax activation. Finally, instantiate the model. The model compiles the model using an Adam optimizer, categorical crossentropy loss, and accuracy metric.

ViT

The model is prebuilt and has the following attributes: 2 classes, learning rate of 0.001, weight decay of 0.0001, batch size of 256, and projection dimensions of 64, and 4 heads. Compile the model with Adam optimizer, categorical crossentropy loss, and accuracy metric (Yin).

Fitting:

The two CNNs using the Keras Sequential API have a batch size of 32, a validation split of 0.2, and 40 epochs. The fitting in the ResNet50 maintained validation split = 0.2 and 40 epochs but the following batch sizes were interchanged: 10, 32, and 60. The fitting in the DenseNet121 model was similarly fitted with 40 epochs and a validation split of 0.2, and the following batch sizes were interchanged: 15, 20, 25, 50. The fitting in ViT maintained the number of epochs but varied the image size by 14, 28, and 48; the batch size was varied by 2, 5, and 7 for an image size of 14, and then the rest of the image sizes had a batch size half of the image size.

Results:

Keras Sequential API CNN

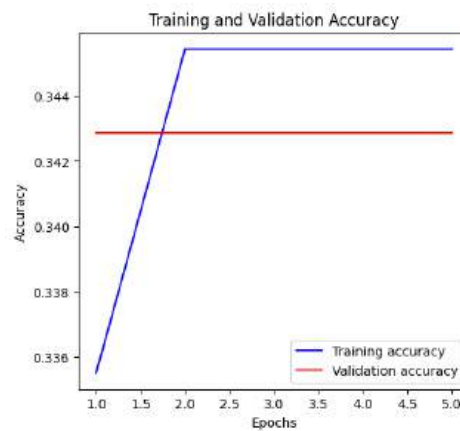
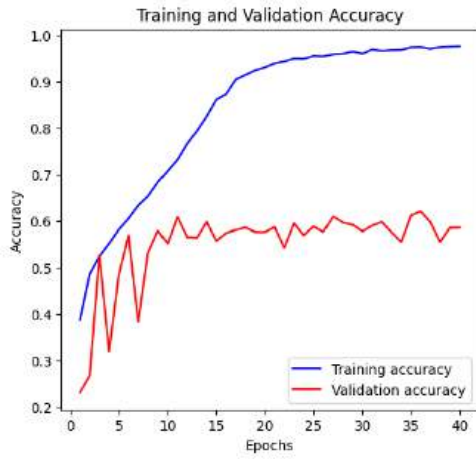


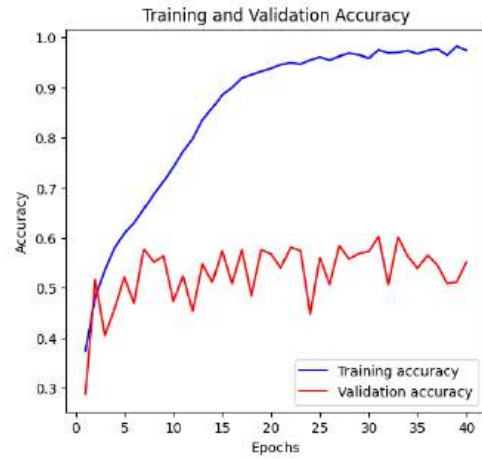
Figure 5. Compares training and validation accuracy of the Keras Sequential API CNN, over a span of 5 epochs

ResNet-50

Batch Size 10



Batch Size 32



Batch Size 60

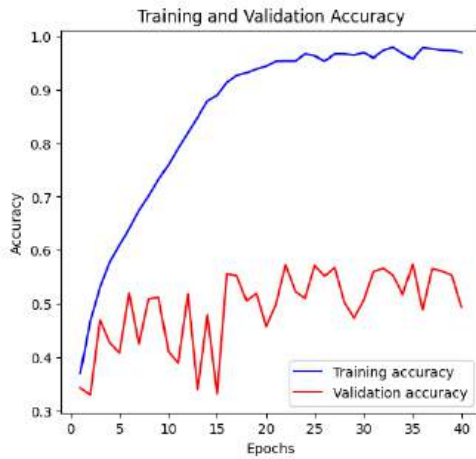
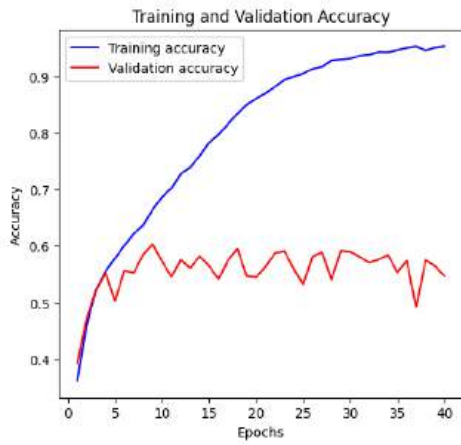


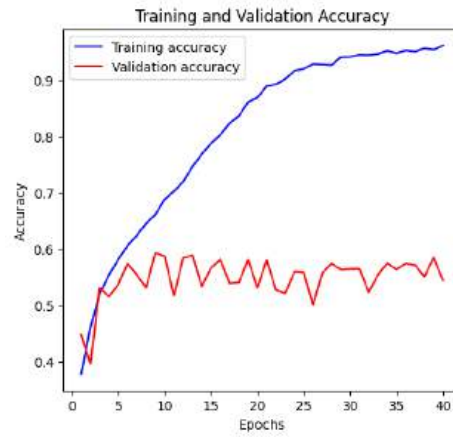
Figure 6. Compares training and validation accuracy of the ResNet-50 model in respect to their corresponding batch size, over a span of 40 epochs

DenseNet121

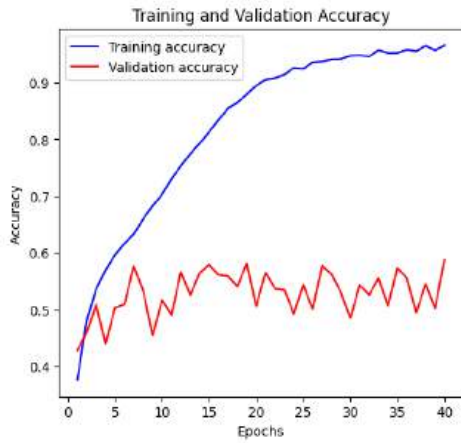
Batch Size 10



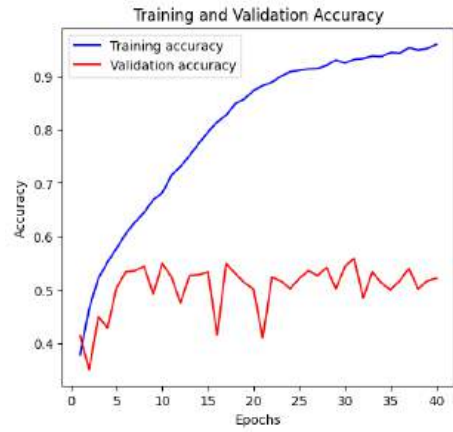
Batch Size 15



Batch Size 20



Batch Size 25



Batch Size 50

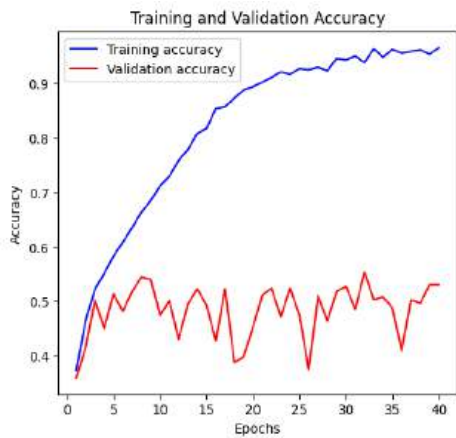


Figure 7. Compares the training and validation accuracy of the DenseNet-121 model in respect to their corresponding batch size, over a span of 40 epochs

ViT

Image size 14, patch size 7

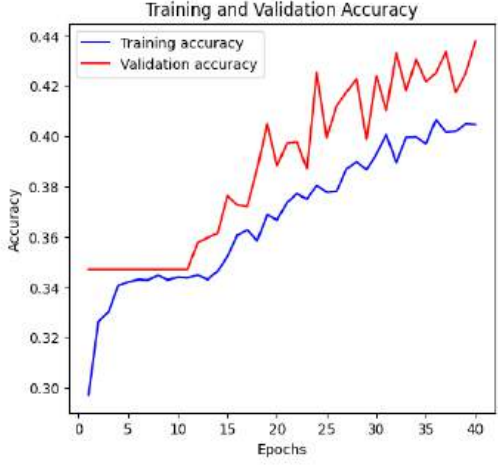


Image size 28, patch size 14

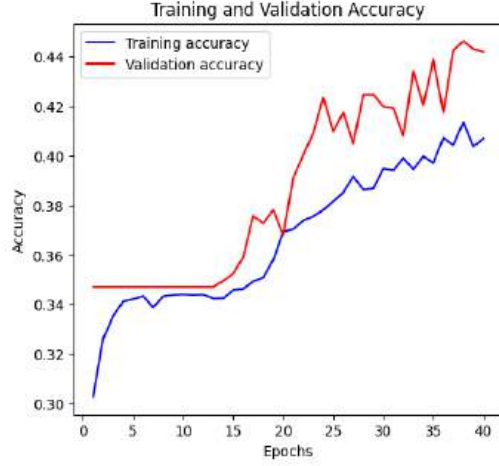


Figure 8. Compares the training and validation accuracy of the ViT model with respect to their corresponding image and patch size, over a span of 40 epochs.

Image size 48, patch size 10

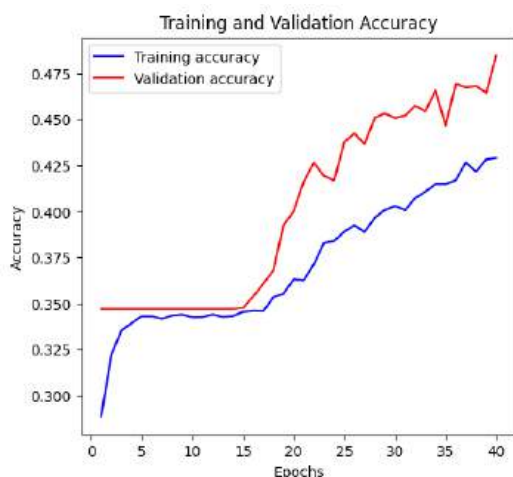


Image size 48, patch size 19

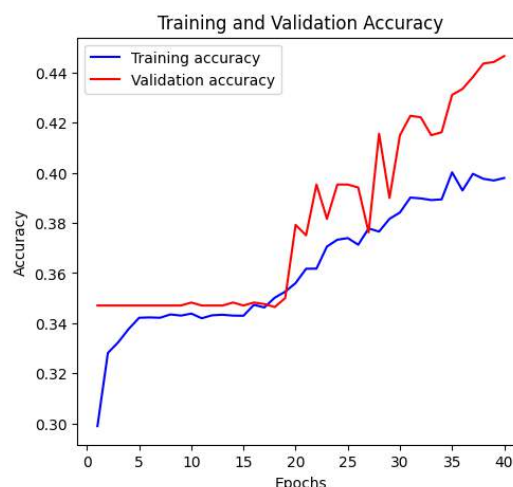


Image size 48, patch size 24

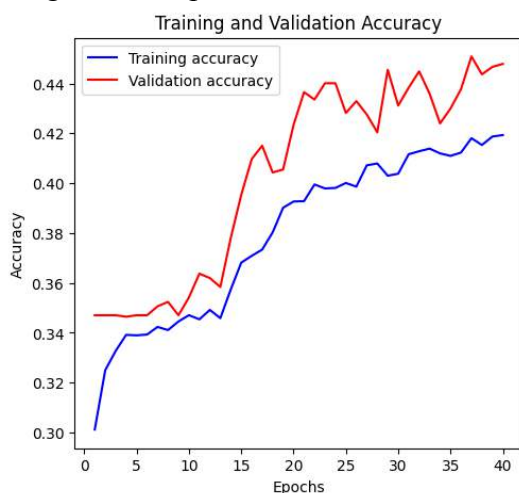


Figure 9. Compares the training and validation accuracy of the ViT model in respect to their corresponding patch size. The image size is maintained 48 over a span of 40 epochs

The CNN using Keras Sequential API yielded a training accuracy of 34.54% and a validation accuracy of 34.21%. This performance was relatively low compared to the other models used, but this may have been due to the use of just 5 epochs.

The ResNet-50 model tested 3 batch sizes: 10, 32, and 60. The batch size of 10 yielded a training accuracy of 97.60% and a validation accuracy of 58.62%. When testing the model with a batch size of 32, it returned a training accuracy of 97.37% and a validation accuracy of 55.16%. With a batch size of 60, ResNet-50 yielded a training accuracy of 96.90% and a validation accuracy of 49.40%. This concludes that as the batch size increases in the ResNet-50 model, both the training and validation accuracies decrease. In addition, the graphs indicating the accuracy over the number of epochs show that the accuracy grows more unstable as the batch size

increases. However, as the batch size decreases, the processing time increases. For example, a batch size of 10 took 30 seconds/epoch, a batch size of 32 took 10 seconds/epoch, and a batch size of 60 took 7 seconds/epoch.

The DenseNet-121 model tests 5 batch sizes: 10, 15, 20, 25, 50. When testing the model with a batch size of 10, it yielded a training accuracy of 95.33% and a validation accuracy of 54.77%. When using a batch size of 15, the model returned a training accuracy of 96.15% and a validation accuracy of 54.47%. DenseNet-121 yielded a training accuracy of 96.53% and a validation accuracy of 58.65% when using a batch size of 20. When using a batch size of 25, it returned a training accuracy of 95.98% and a validation accuracy of 52.14%. A batch size of 50 yielded a training accuracy of 97.20% and a validation accuracy of 57.93%. This concludes that the batch size has a minimal effect on the instantaneous validation accuracy in the DenseNet-121 model. However, the average validation accuracy is greater as the batch size decreases. In addition, both accuracies grow more unstable as the batch size increases. The increased batch time also results in a lower processing time, where a batch size of 10 took 16-17seconds/epoch, a batch size of 15 took 11 seconds/epoch, a batch size of 20 took 8 seconds/epoch, a batch size of 25 took 7-8 seconds/epoch, and batch size of 50 took 4 seconds/epoch.

The ViT tests 3 image sizes: 14, 28, 48. For reference, the original image dimensions for the dataset were 48x48. The image size 14 was tested with a patch size of 7 and yielded a training accuracy of 41.69% and a validation accuracy of 43.77%. The image size 28 was tested with a patch size of 14 and yielded a training accuracy of 43.77% and a validation accuracy of 43.77%. Now when testing the model with an image size of 48, the experiment exchanged patch sizes: 10, 19, 24. The patch size of 10 yielded a training accuracy of 42.90% and a validation accuracy of 48.48%. When testing the model with a patch size of 19, ViT returned a training accuracy of 41.81% and a validation accuracy of 44.66%. In the case of a patch size of 24, it yielded a training accuracy of 43.36% and a validation accuracy of 44.78%. Although the training accuracy has experienced very little growth, the model has yielded a higher validation accuracy as the image size approaches the dataset's original dimensions. In addition, as the patch size grows closer to half the value of the image size, the validation accuracy increases.

Acknowledgement

I'd like to thank my wonderful mentor, Yiqiao Yin, for his encouragement and insightful input. Yiqiao Yin has been essential in improving the approach and breadth of this study.

Conclusion

This paper shows the capabilities of different models like Keras Sequential API CNN, ResNet-50, DenseNet-121, and ViT, in the context of emotion detection. In addition, it demonstrates some of the ways to finetune such a model so that it yields the greatest accuracy. For instance, the paper documents how a lower batch size may increase validation accuracy but may also increase the processing time required per epoch. In addition, the paper demonstrates that patch sizes should ideally be closer to half of the image size. Furthermore, models like

DenseNet-121 and ResNet-50 yield significantly higher accuracies under a lower number of epochs like 40. However, since ViT is a highly complex model and the accuracy graphs highlight steady growth, it is possible that ViT could have performed at a similar, if not better accuracy if the fitting process went through a greater number of epochs. Emotion detection is a growing field and should be studied deeply for society to truly reap the benefits that it has the potential to provide.

Work Cited

- Albahra, Samer, et al. "Artificial Intelligence and Machine Learning Overview in Pathology & Laboratory Medicine: A General Review of Data Preprocessing and Basic Supervised Concepts." *Seminars in Diagnostic Pathology*, vol. 40, no. 2, WB Saunders, 2023.
- Dai, Dengyuhan. "An Introduction of CNN: Models and Training on Neural Network Models." 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR), IEEE, 2021.
- Gajarla, Vasavi, and Aditi Gupta. "Emotion detection and sentiment analysis of images." *Georgia Institute of Technology 1* (2015): 1-4.
- Girasek, Deborah C. "Public Beliefs about the Preventability of Unintentional Injury Deaths." *Accident Analysis & Prevention*, vol. 33, no. 4, 2001, pp. 455-465.
- Han, Kai, et al. "A Survey on Visual Transformer." arXiv preprint arXiv:2012.12556, 2020.
- Jaramillo-Morán, Miguel A., et al. "Improving Artificial Intelligence Forecasting Models Performance with Data Preprocessing: European Union Allowance Prices Case Study." *Energies*, vol. 14, no. 23, 2021, p. 7845.
- Koh, Ezra Y., et al. "Trends in Potentially Preventable Trauma Deaths Between 2005-2006 and 2012–2013." *The American Journal of Surgery*, vol. 218, no. 3, 2019, pp. 501-506.
- Koonce, Brett. "ResNet 50." *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization*, 2021, pp. 63-72.
- Li, Bin. "Facial Expression Recognition by DenseNet-121." *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems*, Academic Press, 2022, pp. 263-276.
- Li, Bin, and Dimas Lima. "Facial Expression Recognition via ResNet-50." *International Journal of Cognitive Computing in Engineering*, vol. 2, 2021, pp. 57-64.
- O'Shea, Keiron, and Ryan Nash. "An Introduction to Convolutional Neural Networks." arXiv preprint arXiv:1511.08458, 2015.
- Pronin, Emily, Daniel Y. Lin, and Lee Ross. "The Bias Blind Spot: Perceptions of Bias in Self Versus Others." *Personality and Social Psychology Bulletin*, vol. 28, no. 3, 2002, pp. 369-381.
- Pronin, Emily, Thomas Gilovich, and Lee Ross. "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others." *Psychological Review*, vol. 111, no. 3, 2004, pp. 781-799.
- Qayyum, Faiza, et al. "A Survey of Datasets, Preprocessing, Modeling Mechanisms, and Simulation Tools Based on AI for Material Analysis and Discovery." *Materials*, vol. 15, no. 4, 2022, p. 1428.
- Ray, Juliet J., et al. "Potentially Preventable Prehospital Deaths from Motor Vehicle Collisions." *Traffic Injury Prevention*, vol. 17, no. 7, 2016, pp. 676-680.
- Rosenthal, Robert, and Kermit L. Fode. "Psychology of the Scientist: V. Three Experiments in Experimenter Bias." *Psychological Reports*, vol. 12, no. 2, 1963, pp. 491-511.

- Sambare, Manos. FER-2013. Kaggle, 2020,
<https://www.kaggle.com/datasets/msambare/fer2013/data>.
- Xu, Yifan, et al. "Transformers in Computational Visual Media: A Survey." *Computational Visual Media*, vol. 8, 2022, pp. 33-62.
- Yin, Yiqiao. Vit.py. GitHub, commit 8da7e7d,
https://github.com/yiqiao-yin/wyn-keras/blob/master/wyn_keras/vit.py.
- Zhu, Yi, and Shawn Newsam. "DenseNet for Dense Flow." 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017.

Is peace in the West Bank and the Gaza Strip possible? By Ganning Ye

Peace in the West Bank and Gaza Strip means peace between Israel and Palestine. They have always shared tense, if not violent relations, with the current conflict one of the most pressing global problems today. At its core, there are deep historical forces, religious beliefs, and complex geopolitical dynamics. This paper analyzes the issue across a range of perspectives, arguing peace is possible, provided the United Nations, the US, and other powerbrokers, form a coalition to influence the respective parties fighting over the West Bank and Gaza Strip. Any solution needs to respect history, reflect modern-day reality, and require painful compromises from all sides.

1. Historical Origins of the Israel-Palestine Issue

Historically, Palestine has been a geographical concept first, and a political or national concept second, as both Jews and Arabs see themselves as descendants of Abraham and the rightful inheritors of his land. From 1020 BCE to 923 BCE, the Jews established the Kingdom of Israel in the region of Palestine (Canaan). In the 1st century BCE, the Roman Empire invaded, causing Jews to disperse around the world. In 622 CE, the Arabs defeated the Roman Empire, and Palestine became part of the Arab Empire. According to the Old Testament of the Bible, Abraham left Mesopotamia and began a new life in Palestine. Jews believe that Abraham is their ancestor, and that Palestine is the promised land given to them by God, a "land flowing with milk and honey." Similarly, Abraham and Hagar's son is considered the ancestor of the Arabs.

From the late 19th century to the early 20th century, the rise of Zionism and the awakening of Arab nationalism has sparked clashes in the region. This stems from the 1920, partition across the Jordan River. The region east of the river is known today as the Kingdom of Jordan, while the region west of the river remained Palestine, a British mandate territory. Currently, most in the international community refers to Palestine as both the West Bank and the Gaza Strip.

Following the events of WW2, the United Nations General Assembly passed Resolution 181 in 1947 to establish a Jewish state and an Arab state in Palestine, with Jerusalem (176 square kilometers) being internationalized. However, this intervention by the international community sparked the first modern-day conflict, with seven member states of the Arab League launching an attack on Israel in May 1948. Since then, major conflicts have occurred between Israel and Arab countries, as well as intermittent skirmishes in response to Israeli settlements.

Despite the international community's attention to the Israel-Palestine conflict and numerous peace initiatives, the irreconcilable core interests of both sides and deep-seated differences in religion, history, and national identity have made the peace process difficult and uncertain. Therefore, the Israel-Palestine issue is not merely a regional conflict but a complex reflection of ethnicity, religion and geopolitics.

2. Current Status of the Israel-Palestine Issue

In 1993, with active mediation by the United Nations and the United States, Israeli Prime Minister Rabin, Foreign Minister Peres, and Palestinian Liberation Organization leader Arafat agreed on the Oslo Accords, establishing a vision of two states coexisting peacefully. Although not fully implemented, it was a phased achievement and major milestone, as The Palestinian Liberation Organization, represented by Fatah, officially abandoned armed struggle in favor of peaceful state-building. In May 1994, the Palestinian National Authority was established, gradually gaining control of the West Bank and Gaza Strip.

However, in June 2007, the radical organization Hamas seized control of the Gaza Strip following a conflict with Fatah, leading to divided governance between the West Bank and the Gaza Strip. Although Israel maintains relative peace with Fatah, and is improving relations with neighboring Arab countries, Hamas continues to advocate armed struggle with a declared intention to "wipe Israel off the map". On October 7, 2023, Hamas launched an attack on Israel, codenamed "Al-Aqsa Flood," prompting Israel to respond with "Operation Iron Sword" against Hamas targets in Gaza. Netanyahu's government has vowed to completely eradicate the Hamas terrorist organization in Gaza through military action, despite significant civilian casualties.

3. Possibility of Achieving Peace Between Israel and Palestine

Under the pressure of the recent Israeli military campaign, Hamas' capabilities have been greatly weakened, leading to the release of some hostages and demonstrating a willingness to compromise through various channels. However, growing civilian casualties threaten to harden attitudes in the region and result in long-lasting ill will. The international community should learn from history; ancient Chinese Confucian thought advocates "harmony is precious" and "harmony without uniformity," while Mohist philosophy promotes "non-aggression." Guided by these principles, Chinese history has seen numerous major integrations of different religions, systems, and ethnicities, providing a successful model for resolving conflicts through peaceful negotiations and political consultations in hotspots around the world." With the collaborative efforts of the United Nations, world and regional powers, and international bodies like the Arab League, achieving peace in the West Bank and Gaza Strip is entirely possible. This can be done through negotiations and consultations that respect historical and current realities, fully address core demands and acknowledge compromises. Israel needs security guarantees, Palestinians need an economic path forward. The following sections will elaborate on this perspective.

3.1 Continued Improvement in the Geopolitical Environment

In 1948, during the outbreak of the First Arab-Israeli War, seven countries in the Arab League formed a coalition to attack Israel. Over time, fewer countries have participated in direct conflict with Israel, with Arab countries gradually accepting the existence of Israel.

For example, in 1974, Egypt and Jordan, having lost territory due to wars with Israel, chose to engage in political negotiations with Israel. Subsequently, with U.S. mediation and assistance, the Camp David Accords led to the peace treaty between Israel and Egypt in March

1979. In October 1994, Jordan signed a peace agreement with Israel, becoming the second Arab country, to establish normal diplomatic relations with Israel. Following this, Israel gradually established formal diplomatic relations with Bahrain, the United Arab Emirates, and Morocco. Under the mediation of the United States, Saudi Arabia, the most influential country in the Middle East and a Sunni Islamic leader, has made breakthrough progress in establishing diplomatic relations with Israel. This provides a model for other countries to normalize relations with Israel. At the same time, Saudi Arabia has also started to normalize relations with Iran, which has deprived Hamas of potential financial and military support. Currently, only Iran, Yemen's Houthi rebels, and Lebanon's Hezbollah maintain a hostile stance towards Israel, with the latter two considered by many to be terrorist organizations. The continuous improvement in the geopolitical environment has reduced pressure on Israel, with most neighboring countries shifting their support for Palestine from military to peaceful state-building. However, this is at risk the longer the current conflict continues.

3.2 Use Historical Examples to overcome Religious and Ethnic Differences

Historically, Islam has been a tolerant religion, maintaining an open attitude towards Catholicism and Judaism, and referring to their adherents as "People of the Book". This tolerance stems from the fact that Islam, in a certain sense, also regards the Bible as a sacred text. The powerful Arab and Ottoman Empires, although Islamic theocracies, displayed varying degrees of kindness and tolerance towards non-Muslims. They even allowed the preservation of certain secular social structures and lifestyles within their territories, including maintaining their own languages, customs, and judicial systems. Jews, for instance, were able to hold high positions such as viziers in these empires without having to convert to Islam, and they were able to maintain their Jewish faith. These policies were effective in the Islamic world at the time, benefiting the governance of the empire and reflecting an Islamic societal attitude of cultural inclusivity.

The rise of Islamic fundamentalism and extremism in the modern era has led to a skewed understanding of Islam's capacity for tolerance. However, the historical tolerance of Islam should not be overshadowed by modern fundamentalism and extremism. Even Fatah, represented by Yasser Arafat, initially engaged in years of violent activities, before developing a more constructive relationship with the Israeli government of Yitzhak Rabin and Shimon Peres. However, in the context of evolving times, both sides set aside past grievances and opted for a moderate approach, achieving phased peace results through negotiation.

3.3 The Importance of Political Wisdom in Promoting Reconciliation

The signing of the Oslo Accords showcased the high political wisdom of leaders like Rabin, Peres, and Arafat. The current war between the far-right government of Israeli Prime Minister Netanyahu and the extremist group Hamas have not achieved their stated outcomes. Over time, public pressure may drive leaders to adopt a more moderate approach. International leadership is needed to promote and support the moderate voices domestically. On June 10,

2024, the United Nations passed a ceasefire agreement between Palestine and Israel drafted by the United States. The main contents include three parts: in the first phase, Hamas and Israel will exchange prisoners for hostages and implement a temporary ceasefire for six weeks; in the second phase, there will be a permanent end to hostilities and Israeli troops will withdraw; in the third phase, reconstruction will begin in Gaza. Hamas has responded positively and has already released some hostages. Israel has also temporarily ceased fire in the Gaza Strip. If both parties can further compromise on this basis, there is a great hope for peace.

3.4 The Role of Economic Development and Education in Regional Peace

The rise of Islamic fundamentalism and terrorism is the result of multiple factors, but economic underdevelopment and low levels of education are key drivers. Islamic fundamentalism refers to a religious movement that insists on returning to the pure initial form of Islam, while terrorism uses illegal violent means to achieve political objectives. Terrorist organizations often exploit social issues like poverty and educational backwardness to recruit members, while also using religious beliefs to indoctrinate and motivate them. Globally, terrorism is not linked to any one religion but is associated with economic backwardness, isolated environments, and low levels of social development and education.

Currently, the relative peace in the West Bank and Israel, is because Fatah has adopted a moderate approach and focused on economic development and peaceful nation-building. Residents can legally enter Israel to work through certain channels, and the per capita GDP has now exceeded 5,000 USD. This has advanced the economy and education in the West Bank, relative to no progress in Gaza. Gaza remains densely populated with low levels of education, economic development, and external aid has become one of the main sources of survival. Hamas has exploited this to recruit new members and gain dominance in Gaza. War has damaged infrastructure, further leading to economic and educational decline.

4. Conclusion

Peace in the West Bank and Gaza Strip still depends on several factors: the establishment of an independent Palestinian state, the Palestinian Liberation Organization and the Palestinian Islamic Resistance Movement bridging their political differences and legally competing within a legal and institutional framework. Neighboring countries should continue to normalize relations with Israel and withhold any support for radical organizations like Hamas, with all parties focusing on developing the economy and education. Although the conflict has not yet ended, in the context of an era focused on peace and development, violence and religious extremism are increasingly unpopular, relations between relevant countries continue to improve, and support for Hamas is dwindling. Through the collective efforts of the international community and sustained negotiations between the involved parties, peace in the West Bank and Gaza can indeed be achieved.

Works Cited

- Ding, Gong. "Why Does the Israel-Palestine Conflict Repeatedly Reignite?" Social Observation, 2014.
- Gong, Zheng. "Reigniting of Conflict Between Israel and Palestine: Will History Repeat Itself?" World Affairs, 2014.
- Gong, Zheng. "Are the Israel-Palestine Peace Talks Heading for Another Suspension?" World Affairs, 2014.
- Global Finance. "History Marred by 'Graffiti'." Global Finance, 2021.
- Montefiore, Simon. Jerusalem: The Biography. Fully Revised and Expanded Edition. Hunan Literature and Art Publishing House, 2019.
- Niu, Song. "U.S. Middle East Peace Policy and Its Future Direction." Asia-Pacific Security and Marine Studies, vol. 1, 2021, pp. 90-103
- Quran
- Tao, Yanzhao. "Gaza, Another 'Ramadan War'." South Wind Window, 2014.
- Xiao, Xian. "Israel-Palestine Relations: Possibly at Rock Bottom, Yet Hope Remains for Peace Talks." World Affairs, 2018.
- Yan, Lishan. "Joyful Journey to Israel." Progressing Together, 2016.

Exploring Motivational Factors in High School Computer Science: A Quantitative Survey Study of Students' Expertise Levels of Java Programming in a Williamson County High School By Krithik Sama

Abstract

This study explores the factors shaping motivation within students enrolled in third-year Java programming courses, aiming to discern the nuance of factors across varying levels of expertise. The research question concerned the motivational factors for high school computer science students and their evolution across expertise tiers, a survey-based approach was employed to amass quantitative data. Contrary to the initial hypothesis of hard work as the predominant motivator, statistical findings unveiled a divergence in motivational drivers across proficiency levels. Novice participants leaned towards extracurricular engagements as stronger factors, while more proficient peers emphasized diligence. The refutation of the hypothesis demonstrates the interactions between motivational factors and proficiency levels. Particularly noteworthy is the marginal role of Java proficiency as a motivator across all expertise tiers. These insights may suggest the development of tailored curriculum interventions aimed at increasing student engagement and motivation within computer science education contexts to better shape learning experiences.

Keywords: motivation, computer science, high school students

Introduction

The field of computer science has undergone constant evolution fueled by technological advancements, exponential growth in computing power, and an increasing desire to embrace innovation. However, the impact of this dynamic sector on its workforce, particularly in terms of motivation—the driving force behind specific actions—remains a pertinent inquiry. Several studies conducted in Brazil have examined the experiences of employees in the computer science domain, spanning from non-profit organizations to major corporations, revealing negative repercussions associated with working in this field. These findings have resulted in a significant downturn in the motivation levels of these workers. Additionally, research in South America has explored the perspectives of university graduates with majors in computer science. Yet, there is a gap in understanding how these findings compare to the experiences of high school students learning computer science. Investigating this aspect is crucial as students represent the future of the field, and it is imperative that they derive maximum benefit from their learning experiences. With these considerations in mind, the research question posed is: "What are the factors influencing motivation among high school students enrolled in a third-year Java programming course, and how do these factors evolve as their expertise levels progress?"

Computer science, centered around software systems, revolves around the framework, development, and application of these systems. Motivation, defined as the impetus for one's actions, plays a pivotal role in this domain (Psychology Today). Despite its abstract nature,

computer science profoundly influences our cognition and emotions. In its nascent stages, workers in computer science were driven by the potential for groundbreaking discoveries at any moment. However, as computer usage proliferated, motivation shifted from passion to financial gain, aligning the profession with higher-paying jobs in society (Carlton). Subsequently, companies began downsizing to enhance productivity and reduce costs, significantly altering the motivational landscape (Lyons). Job security superseded curiosity as the primary motivator, with professionals prioritizing stability and financial well-being. This paradigm shift underscores the profound impact of external factors on the motivational dynamics within the computer science sector. Moreover, the evolving nature of technology continues to shape the motivations of those in the field, highlighting the need for ongoing research into the factors influencing motivation among computer science professionals.

The transition from a pioneering spirit to a pragmatic, survival-oriented mentality highlights the profound socioeconomic ramifications of technological evolution on the workforce. This shift in motivation not only impacted individual practitioners but also reverberated throughout the collaborative and innovative fabric of the field. The once prevalent camaraderie in the early days of computer science, fueled by collective excitement for exploration, gave way to a more competitive and self-preserving ethos (Svitla Systems Inc). Professionals, now driven primarily by the need to safeguard their positions, found themselves compelled to prioritize conformity and efficiency over creative experimentation (Jones and Gottlieb). Amidst this evolving landscape, motivations within computer science mirrored broader societal trends, reflecting the intricate interplay between technological advancement, economic pressures, and personal aspirations. The shift from the thrill of discovery to the pragmatism of job security prompted a reassessment of the intrinsic values guiding professionals in the realm of computer science.

The field of computer science boasts a rich and extensive history, with roots dating back to the innovative concepts of algorithms and sequences pioneered by the Mayan civilization. Prior to the advent of programming languages, the Mayans employed descriptions featuring sequences of calculations on specific data sets, utilizing a base 20 system distinct from the modern base 2 system (Knuth and Pardo). Centuries later, Herman Goldstine and John von Neumann, influential figures in algorithmic science, advocated for a departure from text-based methods in favor of visual representations akin to Mayan approaches. They introduced flowcharts, reminiscent of engineering design processes, to precisely depict algorithms (Knuth and Pardo). These flowcharts comprised four types of boxes: operation, alternative, substitution, and assertion. Operation boxes facilitated transitions in storage, while alternative boxes mirrored conditional statements. Substitution boxes corresponded to variables, and assertion boxes provided guidance for operation boxes. Utilizing these flowcharts, individuals developed intricate data structures like maps and dictionaries. However, contemporary programmers employ iterative segments to access these data structures. One notable example is the "for structure" devised by Rutishauser, a mathematician and computer scientist. While early

programming instruction focused on languages like ALGOL and FORTRAN, Java has emerged as a dominant language in the modern industry (Sobral).

Literature Review

A pivotal study that has informed my research journey examines the motivational factors among software developers in Brazil. Conducted over a decade ago by researchers affiliated with the Center of Informatics at the Federal University of Pernambuco, this study focused on employees of a small software company (Franca et al.). Employing a qualitative research approach centered on interviews, the researchers sought insights into motivation perceptions and their correlation with workplace dynamics. Through categorical analysis of interview transcripts, the team identified self-efficacy as a prominent motivator, particularly underscored by employees' aversion to project delays (Franca et al.). While this study furnishes valuable foundational knowledge for my inquiry, it's pertinent to note that my target participants differ significantly in age. Notably, a related study by the same research team aimed to distinguish between motivation and satisfaction. One significant finding highlighted the importance of collaboration as an indicator of robust motivation among software developers (Franca et al.).

Another research paper that has significantly contributed to my understanding was authored by researchers from a university in Estonia, delving into the factors shaping students' motivation and perception of studying computer science (Säde et al.). The study aimed to develop a scale for quantitatively measuring motivation and perception factors among computer science students, utilizing questionnaires and the Likert scale methodology (Säde et al.). Unlike the qualitative approach adopted in the Brazilian study of 2012, this research employed quantitative methods, offering a complementary perspective on motivational dynamics. While my proposed question shares similarities with this study in exploring factors influencing the study of computer science, it uniquely investigates how expertise levels modulate these factors. It's worth noting the disparity in age demographics between the populations of the two studies, which can potentially yield divergent findings reflective of varied life experiences and educational contexts. By building upon the insights gleaned from both studies, my research endeavors to provide a nuanced understanding of motivational dynamics within the realm of computer science education, accounting for the influence of expertise levels on motivational factors.

An insightful study explores the connection between augmented reality (AR) and motivation in computer science education. Conducted by Buchner and Kerres, the research evaluates AR's effectiveness in enhancing learning abilities and motivation among computer science students. Participants included both first-year students and more experienced individuals in the field. Buchner and Kerres found that while AR benefited first-year students, offering improvements in learning outcomes and motivation, its efficacy diminished among more experienced students. This highlights the nuanced relationship between technological interventions like AR and motivational dynamics in computer science education, emphasizing the importance of considering students' expertise levels.

One reason for selecting this research question is its potential implications. I came across valuable insights on computer science education in a study conducted by researchers at the Department of Mathematics and Computation at the Universidad de La Rioja in Spain. Their research focuses on how a project-based learning (PBL) approach can enhance programming skills (Pérez and Rubio). Participants, consisting of college students, engaged in team-based software tasks mirroring real-world office scenarios (Perez and Rubio). Their findings suggest a positive impact of PBL on programming skills, aligning with the outcomes of the Brazilian study. This likely indicates that factors identified in the Brazilian study are prevalent in PBL, thereby validating their results. However, while the study in Spain examines the efficacy of an educational approach, my research aims to address specific questions regarding motivational factors among high school students in computer science education.

The final article that greatly informed this research project is a German study investigating motivational factors in programming among non-majors in computer science. Conducted by a research team from the Neu-Ulm University of Applied Sciences in Germany, the study sheds light on prevalent motivational trends, particularly highlighting the higher motivation observed among male participants in introductory programming classes (Wolz et al.). However, it's important to note that the participants in this study are slightly older compared to those I intend to survey. This demographic difference could influence the results, providing an opportunity for comparative analysis once the data is gathered.

After reviewing over 30 sources, my hypothesis for the research question is that hard work will be the most prominent factor across the majority of computer science students surveyed in high school.

Methodology

I tailored a methodology to address my research query: "What motivates high school students in a third-year Java programming course, and how does expertise level affect these motivators?" This study focuses on uncovering motivational factors within this specific group and examining how expertise influences these factors. By delving into students' perceptions, I aim to elucidate the key drivers of motivation and gain deeper insights into their experiences in computer science classes. This research has the potential to inform educational strategies and enhance student engagement and learning outcomes in computer science education.

Before formulating my methodology, I conducted a review of existing studies. One such study, conducted by researchers at the Center of Informatics at the Federal University of Pernambuco in Recife, Brazil, aimed to identify the factors of motivation and perceived outcomes of motivation (Franca et al.). This research focused on a group of software developers and employed a maximum variation sampling technique to minimize biases. Data collection regarding motivational factors involved conducting interviews with each participant, prompting them to recount an extraordinarily good day to elicit insights (Franca et al.). Utilizing a qualitative research approach, the study sought to delve deeper into participants' thoughts and experiences, providing rich contextual information.

Another insightful study I encountered focused on investigating motivational factors among non-majors in computer science within introductory classes. Conducted by Wolz and colleagues, this research aimed to elucidate the various motivational factors influencing students not majoring in computer science, with a particular focus on gender analysis. The study participants comprised two cohorts of college freshmen enrolled in introductory programming classes, hailing from two different academic programs (Wolz et al.). Data collection involved administering questionnaires that probed into participants' social backgrounds, ability beliefs, and effort beliefs. Through this comprehensive approach, the study aimed to gain a deeper understanding of the motivational landscape among non-majors in computer science education.

After reviewing the aforementioned studies and additional reports, I proceeded to develop my own research methodology. I considered both qualitative and quantitative research approaches. Qualitative methods focus on obtaining text-based data to explore participants' mental and emotional aspects, while quantitative research utilizes numerical data to identify patterns, relationships, and trends within a larger population. Given the nature of my research question, I opted for a quantitative approach. Although interviews could provide valuable insights, their time-consuming nature for both data collection and analysis posed challenges within the project's timeframe. Thus, I explored alternative methods and turned to surveys, a commonly used tool in qualitative research. Surveys enable the simultaneous collection of extensive information from numerous participants. My objective was to design a survey that primarily gathers quantitative data on students' perceptions of various motivational factors, such as effort, time management, and passion, facilitating subsequent data analysis.

Now that I had chosen how to collect the data, I first had to target a specific group. Since my research question was concerned with students in a third-year computer science course in a high school in Williamson County, I focused on the students who were in those classes. The population size was around 50 students. To gather participants, a consent form was distributed by the computer science teacher. After I received a substantial number of these forms, I then sent out the survey. After waiting for approximately 3 weeks, I closed the survey and started to analyze the data.

The tools I used to perform my method were all Google products, such as Google Forms, Google Docs, and Google Sheets. I made the consent and approval form in Google Docs because that is what's most familiar to me as I use it every day. I used Google Forms to create the survey because that is the only questionnaire creation software that I knew of and had used before. I used Google Sheets to store the data because Google Forms already creates a sheet using the responses given so everything is organized for me.

Results

I ran my survey (see Appendix A) for approximately 2 to 3 weeks, garnering close to 20 responses. The data collected primarily comprised quantitative responses, as the survey employed multiple-choice questions. Each participant's response offered insights into their motivational factors. Upon gathering data via Google Forms and organizing it in a Google Sheet,

I utilized a self-curated translation document to convert multiple-choice responses into numerical values suitable for visual representation. The survey included options for expertise levels, categorized as Beginner (Expertise Level 1), Advanced Beginner (Expertise Level 2), Intermediate (Expertise Level 3), Proficient (Expertise Level 4), and Expert (Expertise Level 5). The tested motivational factors encompassed Java programming, hard work, time management, and extracurricular activities (ECs). These factors were selected based on a synthesis of findings from various research studies.

To illustrate the relationships between expertise levels and motivational factors, I employed a translation table (see Appendix B) to convert multiple-choice responses into numerical values. Subsequently, I calculated the average numerical score for each factor within a given expertise level using the Google Sheets function. Scores ranged from 1 (lowest) to 5 (highest), with higher scores indicating a stronger likelihood of a particular factor serving as motivation for the participant. Using Google Sheets, I generated graphs to visualize trends pertaining to expertise levels and motivational factors. Here are the findings derived from the survey:

Expertise Level	How Influential Is Java	How Influential Is Extracurriculars	How Influential Is Time Management	How Influential Is Hard Work
2	3	5	4	3
2	1	5	4	3
2	4	5	4	4
2	4	5	3	3
2	3	5	3	3
3	3	5	5	5
3	1	5	2	3
3	3	3	4	1
3	2	4	4	4
3	4	4	4	4
3	3	4	3	3
3	3	4	3	2
3	1	1	1	3
3	3	4	4	3
4	4	4	4	4
4	3	4	4	5
5	1	3	5	5

Figure 1: All numerical responses once translated using the document (Appendix B)

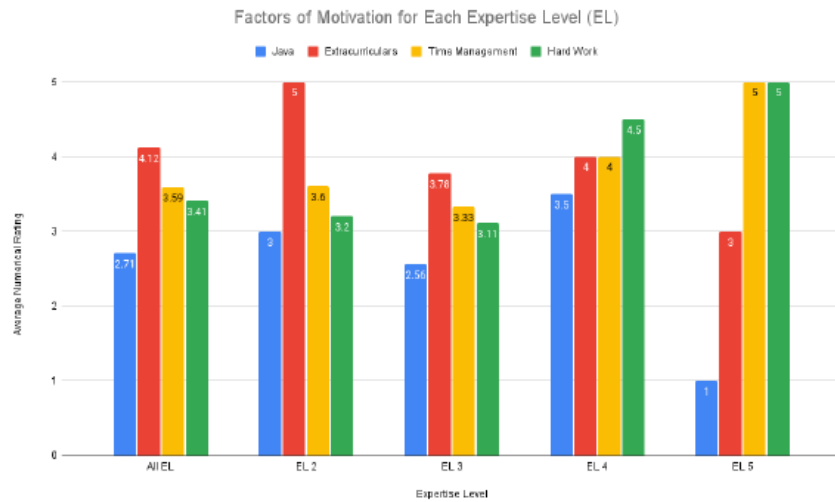


Figure 2: Average scores for each factor for every expertise level (EL)

Figure 2 shows the average score for each motivational factor for each expertise level as well as the combined averages for all expertise levels (leftmost bar stack). This figure allows for the analysis of factors with respect to expertise level and allows for comparison with the average scores for all expertise levels. The highest scoring factor for all expertise levels, EL 2, and EL 3 was extracurriculars. However, for ELs 4 and 5, the highest scoring factor was hard work. This figure allows for generalizations to be made based on the expertise level.

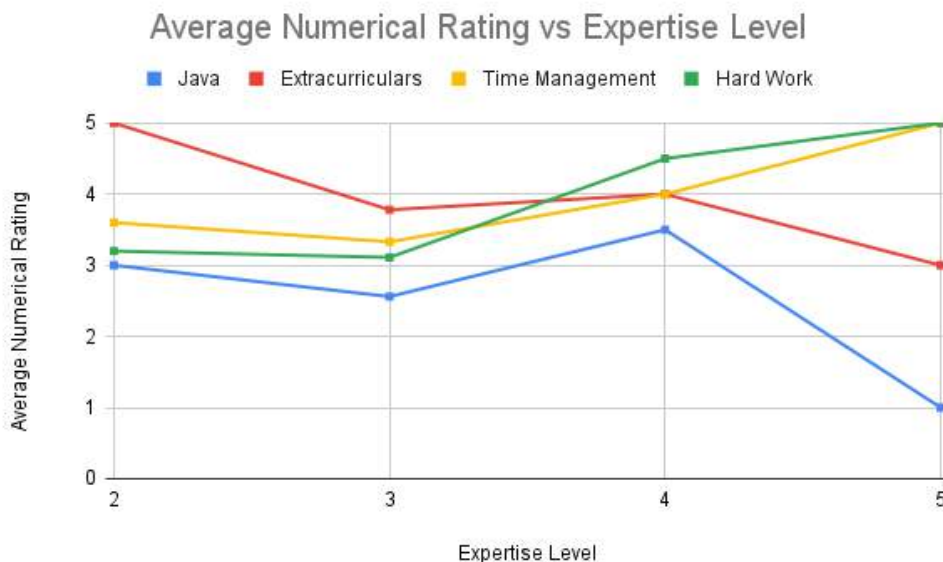


Figure 3: Average numerical rating for each Expertise Level

Figure 3 shows the average numerical rating for each expertise level. This creates different lines where each line represents a separate factor. The independent variable is the expertise level and the dependent variable is the average numerical rating at that expertise level.

The blue and red lines have a downward trend while the yellow and green lines tend to be moving up as the expertise level increases. This figure allows for generalizations to be made based on the factor.

Discussion

According to Figure 2, the most prevalent motivational factor across all participants appears to be extracurricular activities. These activities are perceived as motivational catalysts as they facilitate learning experiences beyond the confines of the classroom, offering valuable opportunities for personal growth. However, a deeper analysis of each expertise level reveals intriguing insights, particularly regarding extracurricular activities. Among participants at the Advanced Beginner and Intermediate levels, extracurricular activities garnered the highest average scores among all motivational factors. It is noteworthy that nearly all participants at these levels expressed aspirations for pursuing careers in STEM fields. This inclination towards STEM careers suggests a desire to pursue higher education post-high school, as many STEM professions require a college degree. Given their perception of their expertise level as relatively low, these participants may seek to enhance their chances of admission to reputable universities by showcasing their involvement in extracurricular activities. Conversely, participants at the Proficient or Expert levels indicated that extracurricular activities were not as significant a motivational factor for them. This divergence in motivational priorities across expertise levels underscores the dynamic nature of motivational factors and their relationship with individual perceptions of skill level and future career aspirations.

Figure 2 also highlights time management as a significant motivational factor across all ELs, with an average score of 3.59. Remarkably, across each expertise level, time management consistently ranked as either the highest or second-highest factor of motivation. Given that all participants in the survey were high school students, the prominence of time management is unsurprising. At this stage, students juggle demanding academic workloads while navigating responsibilities both at home and outside the classroom. Individuals who identified themselves at lower expertise levels may perceive time management as essential for motivation due to the need for additional practice and study in computer science compared to their peers. Conversely, participants at the Proficient or Expert levels may view time management as a means to effectively prioritize areas for improvement, leveraging their existing strengths. Time management serves as a foundational element of motivation by instilling a sense of control and clarity in goal achievement. Moreover, effective time management practices can alleviate stress and prepare students for the demands of the real world. This consistent prominence of time management underscores its enduring significance as a motivational factor in education and beyond.

Conversely, the factor of hard work received an average score of 3.41 when examining all expertise levels. However, upon closer examination of the data within each expertise level, it became evident that the average score for hard work generally increased as the expertise level rose. This trend suggests that individuals with higher levels of expertise recognize the critical

role of hard work in achieving their goals, as the correlation between increased hard work and higher expertise levels is logical. Participants who identified themselves at lower expertise levels tended to attribute higher scores to extracurricular activities, whereas those at higher expertise levels consistently emphasized the importance of hard work. This divergence in motivational priorities may stem from the career aspirations expressed by participants. Specifically, those who indicated Proficient or Expert expertise levels also expressed a desire to pursue careers in computer science or STEM fields, both of which often require college degrees. An intriguing observation is that individuals with higher expertise levels in computer science may already possess a sufficient portfolio of extracurricular activities and thus prioritize hard work as a means to enhance their chances of university acceptance. Conversely, individuals with lower expertise levels in computer science, yet aspiring to pursue STEM careers, may not perceive hard work as a significant motivational factor, potentially due to the belief that minimal effort led to their current expertise levels.

The results regarding the factor of Java programming were particularly surprising, as it consistently emerged as the lowest scoring factor across all expertise levels, despite the participants being enrolled in a computer science class. This unexpected finding may stem from the perception of Java programming as a challenging concept, making it less motivating for students. However, the discrepancy between the importance of Java programming and participants' career aspirations in computer science is notable. Despite almost all of the participants expressing a desire to pursue a career in computer science, Java programming, and programming languages in general, did not emerge as predominant motivational factors. This discrepancy prompts consideration of the possibility that the tested motivational factors are more closely aligned with factors conducive to college admission. This interpretation gains support from the observation that scores for extracurricular activities and hard work were higher compared to other factors within their respective expertise levels. Both extracurricular activities and hard work are typically regarded as essential for college admission, suggesting that participants may prioritize factors perceived to enhance their chances of university acceptance.

Overall, these results and analyses have provided valuable insights and prompted a reevaluation of my research question, offering a new perspective on the complex interplay between motivational factors, career aspirations, and educational goals.

While collecting and analyzing the data, several limitations surfaced. One notable limitation was the significant gender imbalance among participants, with almost all respondents being male. Out of the 17 total participants surveyed, only 2 were female. This skewed gender distribution poses a potential source of bias in the results, as research suggests that males and females may perceive motivational factors differently. Throughout the project, I encountered multiple research studies that delineated variations in motivational factors between men and women. Consequently, the disproportionate representation of genders in the survey sample complicates the process of making generalizations or conducting comprehensive data analysis. The lack of diversity in the participant pool hampers the ability to draw broad conclusions from the findings. Nevertheless, it's important to note that the gender distribution in the survey sample

reflects the composition of the computer science classes at the surveyed school, where the majority of students are male. Despite this alignment with the classroom demographics, the study's findings should be interpreted cautiously and may not be fully representative of the broader population. For future studies, researchers should consider implementing strategies such as offering incentives and employing inclusive survey designs to attract a more diverse range of participants. By addressing these limitations, future research endeavors can enhance the robustness and generalizability of their findings.

During the study, another limitation surfaced regarding the insufficient number of participants for each expertise level. To attain unbiased results, a minimum of 5 participants per expertise level would be optimal given the small population size. However, I encountered challenges in obtaining adequate participation across all expertise levels. Specifically, I received zero participants from the Beginner level, two from the Proficient level, and one from the Expert level. This imbalance in participation led to slightly skewed average scores for each factor, as generalizations were drawn based on the responses of only one or two participants. The scarcity of participants at certain expertise levels hindered the robustness of the findings and limited the extent to which generalizations could be substantiated. Insufficient data impeded the ability to provide conclusive evidence for certain trends or patterns identified in the study. For future research endeavors, it is imperative to strive for a more diverse participant pool to mitigate biases associated with gender and expertise level. While the computer science classes surveyed may have exhibited biases in these aspects, logistical constraints limited access to alternative sources of data. Nonetheless, efforts to expand the participant pool and enhance diversity are crucial for ensuring the validity and generalizability of future research findings.

To mitigate biases, I ensured the survey was anonymous, allowing participants to provide honest responses without fear of repercussion. This anonymity fosters comfort and encourages candid feedback. Additionally, structuring the survey questions as multiple-choice minimizes the potential for misinterpretation or subjective grading. This standardized format enhances consistency and reliability, reducing the likelihood of obtaining false results. These measures aimed to enhance objectivity and reliability while minimizing biases in the study.

Conclusion

The goal of this research study was to determine the motivational factors for computer science students. The research question was “What are the factors of motivation for high school students in a third-year Java programming course and how do these factors change as expertise levels change?” I hypothesized that hard work was the strongest factor of motivation for computer science high school students. To start, I first examined several scholarly sources and studies to determine what information was already available. With this information, I crafted a methodology that would attempt to answer my research question. I implemented the use of surveys that asked about several factors of motivation that I had found in previous research studies. The survey collected quantitative data which allowed me to find trends among motivational factors across four different expertise levels. When analyzing the data, I noticed

many patterns and concluded that my hypothesis was wrong. The answer to my research question is that extracurricular activities were the strongest for lower expertise levels, but for higher expertise levels, hard work was the strongest.

One takeaway I gleaned from this research study is the dynamic nature of the strongest motivational factor across different expertise levels. As expertise level increased, the predominant factor shifted from extracurricular activities to hard work. This shift may be influenced by the students' aspirations for STEM careers, many of which require a college degree. Consequently, the motivational factors assessed in this study may reflect what students perceive as crucial for college applications. However, further research is necessary to substantiate these findings. Another notable finding is that Java programming emerged as the weakest factor of motivation for computer science students, with nearly every participant ranking it with the lowest numerical score. This unexpected result sheds light on the perceived significance of programming language proficiency as a motivational factor within the context of computer science education.

While the study's results are based on a limited scope, drawn from a single high school sample, they offer valuable insights for learning curriculum development. The identification of hard work and extracurricular clubs as the strongest motivational factors suggests opportunities for teachers to enhance student engagement and learning outcomes. Teachers can consider incorporating activities that emulate the dynamics of extracurricular organizations, fostering a more relaxed learning environment conducive to knowledge absorption and experiential learning. Additionally, recognizing that students with higher levels of expertise prioritize hard work, educators can design thought-provoking projects that encourage dedication and effort.

A potential avenue for future research is to explore how different motivational factors impact learning outcomes. This could involve investigating how diverse motivational profiles or experiences influence students' mastery of computer science concepts, performance on assessments, and pursuit of advanced coursework or careers in the field. Furthermore, replicating this study with a larger and more diverse participant group could yield less biased results, offering broader insights into motivational factors among computer science students. Such endeavors can contribute to the refinement of educational practices and the optimization of learning experiences in the field of computer science.

Works Cited

- Buchner, Josef, and Michael Kerres. "Students as Designers of Augmented Reality: Impact on Learning and Motivation in Computer Science." *Multimodal Technologies and Interaction*, vol. 5, no. 41, July 2021, p. 41, <https://doi.org/10.3390/mti5080041>.
- Carlton, Genevieve. "Computer Science: The History of Computer Development." ZDNET, 19 July 2021, www.zdnet.com/education/computers-tech/computer-science-the-history-of-computer-development/.
- França, A. César C., et al. "Motivation of Software Engineers: A Qualitative Case Study of a Research and Development Organisation." *IEEE Xplore*, 1 May 2013, pp. 9–16, <https://doi.org/10.1109/CHASE.2013.6614726>.
- França, A. César C., et al. "Towards an Explanatory Theory of Motivation in Software Engineering: A Qualitative Case Study of a Small Software Company | IEEE Conference Publication | IEEE Xplore." *Ieeexplore.ieee.org*, 2012, ieeexplore.ieee.org/abstract/document/6337859/authors#authors.
- França, César, et al. "Motivated Software Engineers Are Engaged and Focused, While Satisfied Ones Are Happy." *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM '14*, 2014, <https://doi.org/10.1145/2652524.2652545>.
- Jones, Lee Kenneth, and Jessica J. Gottlieb. "Motivational Factors Affecting Computer Science Enrollment." *Proceedings of the ACM Conference on Global Computing Education*, May 2019, <https://doi.org/10.1145/3300115.3312510>.
- Knuth, DONALD E., and LUIS TRABB Pardo. "The Early Development of Programming Languages**the Preparation of This Paper Has Been Supported in Part by National Science Foundation Grant No. MCS 72-03752 A03, by the Office of Naval Research Contract N00014-76-C-0330, and by IBM Corporation. The Authors Wish to Thank the Originators of the Languages Cited for Their Many Helpful Comments on Early Drafts of This Paper.††Reprinted from J. Belzer, A. G. Holzman, and A. Kent (Eds.), 'Encyclopedia of Computer Science and Technology,' Vol. 6, Pp. 419–493. Dekker, New York, 1977. Courtesy of Marcel Dekker, Inc." *ScienceDirect*, edited by N. Metropolis et al., Academic Press, 1 Jan. 1980, pp. 197–273, www.sciencedirect.com/science/article/abs/pii/B9780124916500500198.
- Lyons, Marlo. "5 Reasons People Get Laid Off." *Harvard Business Review*, 23 Jan. 2024, hbr.org/2024/01/5-reasons-people-get-laid-off.
- Pérez, Beatriz, and Ángel L. Rubio. "A Project-Based Learning Approach for Enhancing Learning Skills and Motivation in Software Engineering." *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, Feb. 2020, <https://doi.org/10.1145/3328778.3366891>.
- Psychology Today. "Motivation." *Psychology Today*, 2019, www.psychologytoday.com/us/basics/motivation.
- Säde, Merilin, et al. "Factors That Influence Students' Motivation and Perception of Studying Computer Science." *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, Feb. 2019, <https://doi.org/10.1145/3287324.3287395>.

Sobral, Sónia Rolland. “30 Years of CS1: Programming Languages Evolution.” Uportu.pt, June 2020, <http://hdl.handle.net/11328/2925>.

Svitla Systems, Inc. “What Causes a Lack of Motivation in Software Developers and How to Prevent It.” Svitla.com, Svitla Systems, Inc, 26 Feb. 2018, svitla.com/blog/what-causes-a-lack-of-motivation-in-software-developers-and-how-to-prevent-it.

Wolz, Sabine, et al. “Influence Factors on Students Motivation in Introductory Programming Lectures of Computer Science Non-Majors.” Cogent Education, edited by Taralynn Hartsell, vol. 9, no. 1, Mar. 2022, <https://doi.org/10.1080/2331186x.2022.2054914>.

The Efficacy of AI Voice Agents in Enhancing Mental Health Outcomes: A Preliminary Investigation By Dhruv Reddy, Rithvik Rathinsabapathy

Abstract

This preliminary study investigates the short-term efficacy of an AI voice agent in improving mental health outcomes, focusing on symptoms of anxiety and depression. Conducted over two months with 20 participants, the study integrates quantitative measures, such as the GAD-7 (Spitzer, Kroenke and Williams) and PHQ-9 scales (Kroenke, Kurt and Spitzer), with qualitative data from user interviews and satisfaction surveys. The AI voice agent, developed using ReactJS for the front end and Python for the back end, incorporates a speech-to-text pipeline, a fine-tuned GPT-3.5 model (Model X) for the natural language processing (NLP) component, and a text-to-speech module. The core NLP (Clark, Fox and Lappin) model was embedded and fine-tuned to cater specifically to mental health dialogue. The system utilizes APIs like VapiAI (vapi) for speech-to-text conversion, low-latency audio streaming, optimizations, and more to enhance user experience. The findings indicate notable reductions in anxiety and depression symptoms, suggesting that AI voice agents can effectively supplement traditional mental health care. The study also discusses ethical considerations, emphasizing data privacy, algorithmic transparency, and user interaction design.

1. Introduction

1.1 Background and Significance

Mental health disorders, including anxiety and depression, are significant contributors to global morbidity and disability. According to the World Health Organization, depression affects over 280 million people worldwide, leading to substantial disability and economic burden (WHO). Traditional mental health services, including psychotherapy and pharmacotherapy, often face barriers such as high costs, limited availability, and societal stigma, which hinder access to care and lead to untreated or inadequately treated conditions.

The integration of artificial intelligence (AI) technologies in healthcare has introduced new opportunities for delivering mental health support. AI voice agents, which enable real-time, conversational interactions, have the potential to expand access to mental health services, offering a scalable and flexible solution. Unlike text-based systems, voice agents can engage users in more natural and empathetic dialogues, potentially enhancing the therapeutic experience (University of Nevada). This study assesses the effectiveness of an AI voice agent in reducing symptoms of anxiety and depression, leveraging advanced technology and clinically validated therapeutic content.

1.2 Research Objectives

The primary objective of this research is to evaluate the short-term effectiveness of an AI voice agent in reducing symptoms of anxiety and depression. This involves measuring changes

in participant-reported symptoms using standardized psychological scales and assessing the clinical significance of these changes. Additionally, the study aims to explore user engagement and satisfaction with the AI voice agent, focusing on ease of use, perceived empathy, and overall interaction quality. The study also seeks to examine ethical issues related to AI in mental health care, particularly concerning data privacy, informed consent, and algorithmic bias (D. D and Zokaei). By addressing these objectives, the study aims to establish guidelines for the responsible deployment of AI technologies in healthcare settings, ensuring equitable benefits for all users.

2. Literature Review

2.1 AI in Mental Health Care

The application of AI in mental health care has gained considerable attention, with innovations ranging from diagnostic tools to therapeutic interventions. AI-based systems, including chatbots and voice agents, use natural language processing (NLP) and machine learning algorithms to interact with users, providing support and guidance based on the analysis of textual and vocal data. Early applications focused on chatbots designed to deliver cognitive-behavioral therapy (CBT), psychoeducation, and self-management strategies (G, Wang and Graham). Studies have shown that these AI systems can reduce symptoms of depression and anxiety, offering a cost-effective and accessible complement to traditional therapy. For example, Fitzpatrick et al. (2017) found that a CBT-based chatbot reduced anxiety and depression symptoms over two weeks (Fitzpatrick, Kathleen and Darcy).

However, text-based interactions often lack the emotional depth and nuance necessary for effective therapeutic engagement. This limitation has led to increased interest in voice-based AI technologies, which can convey tone, emotion, and other non-verbal cues critical for meaningful communication. Voice agents can simulate human-like interactions, providing users with a more immersive and empathetic experience. The ability to detect and respond to vocal cues, such as changes in pitch and pace, enhances the agent's capacity to deliver personalized support, potentially leading to improved mental health outcomes.

2.2 The Unique Potential of Voice Agents

Voice agents, powered by advanced NLP and machine learning, can engage users in dynamic, responsive conversations. This capability is particularly valuable in mental health care, where the therapeutic alliance – a collaborative and empathetic relationship between therapist and patient—is a key factor in treatment success. Voice agents can recognize and interpret a range of vocal cues, adjusting their responses to match the user's emotional state. This feature allows the agent to provide real-time feedback and support, enhancing the user's experience of empathy and understanding.

3. Methodology

3.1 Participants

The study recruited 20 participants through online mental health forums, social media platforms, and local community centers. Inclusion criteria required participants to be adults aged 18-65 who scored ≥ 10 on the GAD-7 or ≥ 10 on the PHQ-9, indicating at least moderate levels of anxiety or depression. Exclusion criteria included current engagement in psychotherapy, use of psychotropic medications (Wikipedia) initiated within the past month, or diagnosis of severe mental illness such as schizophrenia or bipolar disorder. The study aimed for diversity in the sample, including variations in age, gender, socio-economic status, and cultural backgrounds.

3.2 Study Design

The study utilized a pre-post quasi-experimental design (Scribbr) over an eight-week intervention period. Participants interacted with the AI voice agent three times per week, with each session lasting 15-30 minutes. This schedule was designed to provide more frequent support than traditional weekly therapy sessions, allowing for a more immersive experience with the AI agent. Baseline assessments of anxiety and depression symptoms were conducted using the GAD-7 and PHQ-9 scales, followed by assessments at four weeks (mid-intervention) and eight weeks (post-intervention). This design allowed for tracking of participant progress at regular intervals throughout the study period, providing insights into the rate and pattern of symptom changes.

3.3 Instruments and Measures

Quantitative data were collected using the Generalized Anxiety Disorder 7-item scale (GAD-7), which measures the severity of anxiety symptoms, with a score range from 0 to 21. This tool is widely used for its reliability and validity. Additionally, the Patient Health Questionnaire-9 (PHQ-9) was used to assess the severity of depressive symptoms, with scores ranging from 0 to 27. The PHQ-9 is valued for its ease of use and strong psychometric properties.

Qualitative data was gathered through semi-structured interviews and satisfaction surveys conducted post-intervention. These interviews explored participants' experiences with the AI voice agent, including perceived effectiveness, emotional support, usability, and challenges encountered. The qualitative data provided rich, contextual insights into the participants' subjective experiences, complementing the quantitative findings.

3.4 AI Voice Agent Description

The AI voice agent was developed using ReactJS for the frontend and Python for the backend, incorporating a sophisticated pipeline for speech-to-text conversion, natural language processing (NLP) with GPT-3.5, and text-to-speech output. While the core NLP component, referred to as Model X, was fine-tuned and embedded for specific mental health dialogue, the system also utilized APIs like VapiAI for other critical functions. These included speech-to-text

conversion, low-latency audio streaming, and additional optimizations to enhance the overall user experience.

The fine-tuning and prompting of GPT-3.5 was conducted to align the model with mental health scenarios, using a dataset enriched with over 200 lines of open-source therapy simulations focused on anxiety and depression. These simulations were supplemented with content from academic textbooks on cognitive-behavioral therapy, mindfulness, and emotional regulation. To ensure the AI's behavior aligned with therapeutic standards, a detailed system prompt was implemented, instructing the AI to act as an empathetic and non-judgmental mental health therapist, with emphasis on confidentiality, support, and evidence-based practice.

The system featured technical optimizations, including GPU-accelerated inference for faster processing and intelligent caching to minimize latency (Jiang and Zihan). These enhancements were particularly important in creating a seamless and responsive interaction experience. The AI voice agent also incorporated a pause detection mechanism, allowing the agent to recognize natural pauses in conversation and stop speaking in response to user interruptions. Such were primarily done through VAPI and other APIs. Additionally, as shown in figure 1 to the left, the interface design was carefully crafted to replicate the nuances of traditional therapy sessions, offering intuitive controls such as mute and end call options, and utilizing a soothing female voice sourced from the ElevenLabs API (Elevenlabs). This design aimed to foster a supportive and user-centered experience, encouraging openness and trust.

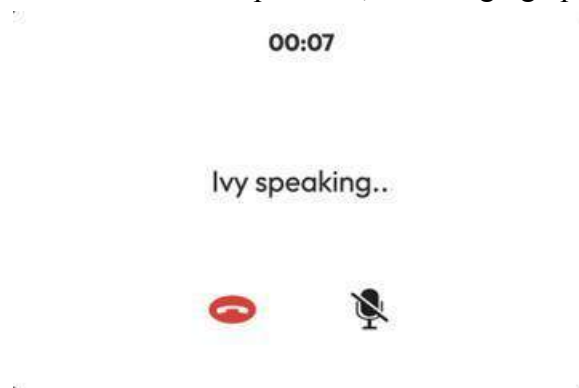


Fig 1: User interface

4. Results

4.1 Quantitative Findings

The analysis of the GAD-7 and PHQ-9 scores revealed notable reductions in anxiety and depression symptoms among participants. The mean GAD-7 score decreased from 13.5 at baseline, indicating moderate anxiety, to 10.5 post-intervention, indicating mild anxiety. This reduction was statistically significant ($p < 0.05$), with an effect size (Cohen's d) (Becker) of 0.6, suggesting a moderate practical impact. Similarly, the mean PHQ-9 score decreased from 15.2, representing moderate depression, to 12.3, representing mild depression, also statistically significant ($p < 0.05$), with a Cohen's d of 0.5, indicating a moderate effect.

Measure	Pre-Intervention	Post-Intervention	Change	p-Value
GAD-7 Score	13.5	10.5	-3.0	≤0.05
PHQ-9 Score	15.2	12.3	-2.9	≤0.05

Table 1: Mean GAD-7 and PHQ-9 scores Before and After Intervention

Table 1 illustrates the reductions in GAD-7 and PHQ-9 scores, highlighting the notable improvements in anxiety and depression symptoms post-intervention.

4.2 Qualitative Findings

Qualitative data from semi-structured interviews and satisfaction surveys provided deeper insights into participants' experiences with the AI voice agent. Most participants reported high satisfaction levels, emphasizing the system's accessibility, ease of use, and empathetic interactions. Users particularly appreciated the immediate availability of the agent and its ability to engage in meaningful conversations, which they found comforting and supportive. Participants noted that the voice agent's empathetic responses and ability to understand and reflect on emotional cues significantly enhanced their experience. However, some users mentioned challenges with the agent's responses during more complex emotional interactions or when nuanced emotional expressions were involved. These limitations were attributed to the current state of NLP technology, which may not fully capture the depth of human emotions.

Context	Summary
Emotional Support	The agent's empathetic responses were frequently praised.
Interaction Quality	High satisfaction with conversation quality, though some limitations in understanding complex emotions were noted.
Usability	Generally user-friendly, with intuitive design and operation.
Areas for Improvement	Need for enhanced emotional intelligence and handling of complex emotional states.

Table 2: Summary of Qualitative Feedback

5. Discussion

5.1 Efficacy of AI Voice Agents

The results from this study suggest that AI voice agents can significantly reduce symptoms of anxiety and depression in a short-term intervention. The observed reductions in

GAD-7 and PHQ-9 scores indicate that participants experienced notable improvements in their mental health, likely due to the consistent and responsive interaction provided by the AI voice agent. The use of advanced NLP and machine learning techniques allowed the agent to engage users in meaningful, personalized conversations. This level of personalization is crucial in mental health care, where the effectiveness of interventions often depends on the perceived empathy and relevance of the support provided.

The AI voice agent's ability to deliver real-time, individualized responses plays a key role in its therapeutic potential. By tailoring its interactions based on user input, the agent can offer specific coping strategies, provide emotional validation, and reinforce positive behaviors. This level of customization helps build a therapeutic alliance, an essential component of effective mental health treatment, fostering a sense of understanding and support.

5.2 Ethical and Practical Implications

The integration of AI into mental health care introduces significant ethical and practical challenges that necessitate careful consideration. A primary concern is the protection of sensitive mental health data. Implementing comprehensive data protection protocols, such as robust encryption and secure storage, is essential to safeguard user information (Shojaei, Parisasadat and Vlahu-Gjorgievska). Transparency in these practices is crucial; users must be fully informed about how their data is collected, utilized, and protected. This transparency helps to build trust and fosters a secure environment for sharing personal information.

Informed consent is another critical issue. It is imperative to communicate clearly with users about the capabilities, limitations, and potential risks of AI systems. Users should have a thorough understanding of the nature of AI interactions, including the scope of data collection and analysis. Such transparency is vital in empowering users to make informed decisions about their engagement with technology, thereby respecting their autonomy and rights.

Moreover, the issue of algorithmic bias presents a substantial ethical challenge (Gaonkar, Cook and Macyszyn). AI systems are at risk of perpetuating biases inherent in their training data, which can result in unequal treatment outcomes. This concern is particularly acute in mental health care, where biases can disproportionately affect marginalized and vulnerable populations (Hunter and Erickson). Addressing this requires the implementation of rigorous bias detection and mitigation strategies. Regular audits and continuous monitoring of AI performance are necessary to identify and correct biases, ensuring equitable treatment for all users. Additionally, it is critical to curate diverse and representative training datasets to enhance the fairness and inclusivity of AI applications in mental health care.

By addressing these ethical implications, the responsible deployment of AI in mental health care can be achieved, ensuring that the technology not only enhances mental health outcomes but also adheres to the highest ethical standards.

5.3 Limitations and Future Research

This study has several limitations that should be considered when interpreting the findings. One primary limitation is the small sample size, which restricts the generalizability of the results. The short duration of the intervention (two months) also limits the ability to conclude about the long-term efficacy of AI voice agents. Future studies should include larger, more diverse participant samples and longer follow-up periods to better assess the sustainability and broader applicability of the observed benefits.

Additionally, technical limitations related to the AI voice agent's implementation, such as latency and response times, can affect user experience and the perceived effectiveness of the intervention. Despite optimizations like GPU-accelerated inference and intelligent caching, some participants may experience delays or interruptions that could diminish the therapeutic value of the interaction. These challenges underscore the importance of ongoing development and refinement of AI technologies to improve their performance and reliability in real-world settings.

Another area for future research is enhancing the AI voice agent's emotional intelligence. While the current system can provide basic empathetic responses, it may struggle with more complex or nuanced emotional expressions. Integrating multimodal data sources, such as facial expressions, voice tonality, and physiological signals, could enhance the agent's ability to interpret and respond to users' emotional states accurately (Dhanith and Joe). Such advancements would likely improve the overall user experience and increase the system's therapeutic effectiveness.

6. Conclusion

The findings from this preliminary study suggest that AI voice agents can serve as an effective adjunct to traditional mental health services, offering a scalable and accessible solution for individuals experiencing anxiety and depression. The reductions in GAD-7 and PHQ-9 scores indicate that such technologies have the potential to improve mental health outcomes meaningfully. However, the deployment of AI in mental health care must be guided by rigorous ethical standards, ensuring data privacy, transparency, and equity. As AI technologies continue to evolve, ongoing research and development will be essential to optimize these systems for mental health applications. This includes enhancing the emotional and conversational capabilities of AI voice agents and addressing the ethical challenges associated with their use. By prioritizing these areas, the field can harness the full potential of AI to expand access to mental health care and support the well-being of individuals worldwide.

Acknowledgments

We would like to express our sincere gratitude to Arjun Reddy, Yana Thakkar, and Riya Shankar for their invaluable contributions to this research. Arjun, Yana, and Riya assisted with research and data collection, providing essential support in gathering and organizing the data. Their efforts significantly contributed to the depth and quality of this study. We also acknowledge the 20 participants in the study for their time and openness, which made this research possible.

Works Cited

- D. D, Farhud and S Zokaei. "Ethical Issues of Artificial Intelligence in Medicine and Healthcare." Iranian Journal of Public Health 50.11 (2021): i-v. <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8826344/>>.
- Becker, Lee A. n.d. <<https://www.uv.es/friasnav/EffectSizeBecker.pdf>>.
- Clark , Alexander , Chris Fox and Shalom Lappin. The Handbook of Computational Linguistics and Natural Language Processing. Blackwell Publishing Ltd, 2010. <<https://onlinelibrary.wiley.com/doi/book/10.1002/9781444324044#page=291>>.
- Dhanith and P. R Joe. "Multimodal Emotion Recognition Using Audio-Video Transformer Fusion with Cross Attention." arXiv (2024). <<https://arxiv.org/abs/2407.18552>>.
- Elevenlabs. n.d. <<https://elevenlabs.io/api>>.
- Fitzpatrick, et al. "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial." JMIR Mental Health 4.2 (n.d.). <<https://mental.jmir.org/2017/2/e19>>.
- G, Nicol, et al. "Chatbot-Delivered Cognitive Behavioral Therapy in Adolescents With Depression and Anxiety During the COVID-19 Pandemic: Feasibility and Acceptability Study." MIR Formative Research 6.11 (2022). <<https://formative.jmir.org/2022/11/E40242/>>.
- Gaonkar, B, K Cook and L Macyszyn. "Ethical Issues Arising Due to Bias in Training A.I. Algorithms in Healthcare and Data Sharing as a Potential Solution." The AI Ethics Journal 1.1 (2023). <<https://aiej.org/aiej/article/view/1>>.
- Hunter, Kevin and Jennifer S Erickson. "Addressing Inequities in Access to Mental Healthcare: A Policy Analysis of Community Mental Health Systems Serving Minoritized Populations in North Carolina." Administration and Policy in Mental Health and Mental Health Services Research 50.1 (2023): 68-82. <<https://link.springer.com/article/10.1007/s10488-024-01344-8>>.
- Jiang and Zihan. " Systematic Methodology for Analysis of Deep Learning Hardware and Software Platforms." CCF Transactions on High Performance Computing (2021). <https://proceedings.mlsys.org/paper_files/paper/2020/file/650e2245aa3513ed517f4cf1b3d58e06-Paper.pdf>.
- Kroenke, et al. "The PHQ-9: Validity of a Brief Depression Severity Measure." Journal of General Internal Medicine 16.9 (2001): 606–613. <<https://onlinelibrary.wiley.com/doi/full/10.1046/j.1525-1497.2001.016009606.x>>.
- Scribbr. n.d. <<https://www.scribbr.com/methodology/quasi-experimental-design/>>.
- Shojaei, Parisasadat and Elena Vlahu-Gjorgievska. "Security and Privacy of Technologies in Health Information Systems: A Systematic Literature Review." Computers 13.2 (n.d.): 41-56. <<https://www.mdpi.com/2073-431X/13/2/41>>.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. n.d.
- Spitzer, Robert L, et al. "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7." Archives of Internal Medicine 166.10 (2006): 1092–1097. <<https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/410326>>.
- University of Nevada, Las Vegas,. 2022. <Hasanzadeh Vayghan, Saeed>.
- vapi. n.d. <<https://vapi.ai/>>.
- WHO. n.d. <<https://www.who.int/news-room/fact-sheets/detail/depression>>.
- Wikipedia. n.d. <https://en.wikipedia.org/wiki/Psychoactive_drug>.

Establish Data Protection To Ensure Privacy In the Surveillance Era By Audrey Horn

Abstract

The United States government's secretive mass surveillance of the public garnered global attention following the revelations by Edward Snowden. This surveillance infringed upon the privacy of citizens and highlighted the need to reform U.S. digital surveillance and privacy policies. To answer the critical question of what steps should be taken to protect privacy while addressing the security needs, this study delves into the full extent of surveillance by the National Security Agency. It also examines methods used to anonymize and pseudonymize individuals online. This study finds that current anonymization and pseudonymization techniques are inadequate and can be easily bypassed. Consequently, this study recommends that the U.S. should establish a foundational privacy law with concrete rules on surveillance. The General Data Protection Regulation is identified as a potential model for the U.S. government to consider when implementing foundational privacy regulations. Should the U.S. government choose to do so, the enactment of such privacy regulations will be instrumental in ensuring the right to privacy for its citizens.

Introduction

In 2013, National Security Agency (NSA) contractor Edward Snowden released evidence of mass surveillance of the public by the NSA in the U.S. and the Government Communications Headquarters (GCHQ) in Britain (Greenwald, 2013). Snowden's revelations prompted Brazilian President Dilma Rousseff to berate the NSA and GCHQ practices, stating that "in the absence of the right to privacy, there can be no true freedom of expression and opinion and therefore no effective democracy" (Wong, 2015). Article 12 of the Universal Declaration of Human Rights (UDHR) protects the right to privacy; however, both the NSA and GCHQ have disregarded this right, as shown by their previously undisclosed mass surveillance (Kampmark, 2014). The two aforementioned surveillance organizations justified their actions by claiming that they were only storing metadata, which is defined as data that provides information about other data (Feldstein, 2019; Bauman et al., 2014). In agreement with the NSA and the GCHQ, the Obama administration argued that storing metadata is different from intrusive surveillance operations (Kampmark, 2014). However, the general public contends that such secret surveillance is unjust and illegal (Austin, 2015). The UN High Commissioner and Human Rights Watch appealed to the public, advocating for protection against these types of surveillance agencies (Kampmark, 2014). To counter the invasive nature of mass surveillance, laws and regulations should be updated to clearly define legitimate versus illegitimate surveillance, taking into account public opinions, technology advancements, and national security. The issue of mass surveillance has raised questions about whether U.S. citizens truly have the right to privacy as stated in the Fourth Amendment of the U.S. Bill of Rights. A comprehensive nationwide law similar to the European Union's (EU) General Data Protection Regulation (GDPR) should be implemented to increase the transparency of the United States government's methods of data collection and ensure the

recognition of individual privacy. To reach this conclusion, the following discussion will provide a background on the issue of surveillance and present a review of current solutions.

Background

The issue of privacy and security was brought to light by Edward Snowden, who faced legal consequences after leaking documents discussing the mass surveillance of phone calls and internet inquiries by U.S. surveillance agencies to the media. Despite knowing the risk of severe prosecution, Snowden stated that “[he did] not want to live in a world where everything [he did] and [said was] recorded” (Chaleff & Thomas, 2017). In the end, Snowden’s whistleblowing sparked public interest and awareness regarding surveillance practices, leading to the formation of groups dedicated to investigate such practices, exemplified by Germany’s Bundestag Investigation Committee (Colvin, 2018). These communities have demonstrated considerable influence in shaping surveillance policies, which is significant in light of the breach of the general public’s privacy. Ultimately, the findings of these investigations led to the designation of metadata collection as a state crime internationally, leading to the UN resolutions and the appointment of a Special Rapporteur on the Right to Privacy (Colvin, 2018). Through Snowden’s revelation of surveillance by the U.S. agencies, it became apparent that a secret NSA program, PRISM, had engaged in large-scale mass surveillance of governments, companies, and citizens, including those in some of the U.S.’s closest allies in Europe and Latin America. The PRISM program was able to obtain metadata by data harvesting through internet cables, enabling the government to track all searches conducted by individuals (Bauman et al., 2014). Additionally, the British surveillance agency, the Government Communications Headquarters (GCHQ), was also found to be collaborating with the NSA in exploiting metadata for intelligence purposes (Kampmark, 2014). Furthermore, private companies such as Google, Microsoft, Apple, and Skype regularly collect data for commercial purposes and provide such data to intelligence services like the NSA without user permission. Data is also gathered through platforms like Microsoft and Dropbox, serving as a reminder to the public that no online data is truly private (Bauman et al., 2014). These data were used to map relationships between people globally, disclosing their IP addresses, personal information, and personal interests to the U.S. and U.K. governments.

The government, however, was determined to prove that this secret surveillance was justified and that Snowden was the one who compromised security. U.S. intelligence officials reported that Snowden’s whistleblowing led to the endangerment of U.S. personnel and facilities globally, damaged intelligence efforts, and undermined partnerships between the U.S. government and other nations (Coyne, 2019). For example, Snowden’s revelation exposed that the NSA intentionally tapped German Chancellor Angela Merkel’s devices, causing widespread shock among the EU and demands for reformation. As mentioned before, the Obama administration argued that storing metadata differed from unlawful surveillance (Kampmark, 2014). Even the NSA itself claims that everything that they pursue is just and fully complies with all laws (Austin, 2015). In summary, the NSA and the U.S. government appear to be defensive of

their actions, portraying their actions as just. However, the general public feels that the opposite is true; many protest against surveillance, as they believe that it is completely unjust and illegal (Austin, 2015). Rooted in general human nature, individuals do not want their privacy infringed upon secretly — it feels like a violation of human rights.

Overall, there are many conflicts between the government and surveillance agencies on one side, and the general public and the UN on the other. While the government seeks to obtain metadata from the public, individuals are opposed to having their private information disclosed to unfamiliar government officials. Policymakers face the daunting challenge of finding the balance between privacy regulations that protect individual freedoms and the need for effective crime and terrorist investigations facilitated by metadata collection (Hagen & Lysne, 2016). For various reasons, the government may feel compelled to prioritize either privacy or data collection, making it very difficult to develop a just policy to satisfy both the public and the government. Nevertheless, some solutions have been developed to attempt to reconcile these two objectives.

Existing Methods for Data Protection

Personal data protection is foundational to privacy rights. In the United States, personal data encompasses information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information that is linked or linkable to a specific individual (NIST, n.d.). Various methods are available to protect the privacy rights of individual data subjects, including data anonymization and pseudonymization. These methods continue to evolve due to their wide impact on privacy and security.

Data anonymization renders the identification of the data subject impossible in an irreversible way. Anonymization can be done by various methods, such as randomization and generalization. However, anonymization may not always be possible due to specific circumstances related to the nature, storage, or use of data, among other reasons. In such cases, pseudonymization may be used to protect the privacy of data subjects. GDPR Article 4 defines pseudonymization as “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.” Pseudonymization replaces any information that could potentially identify an individual with another symbol or word (Data Protection Commission, n.d.). By preventing the direct link of associating specific information to an individual, pseudonymization addresses the concern about personal information leakage and fulfills part of the public's wish for privacy. However, the data subject may be re-identifiable by using additional information (European Data Protection Supervisor, n.d.). This additional information could come from Social Network Analysis (SNA), the analysis of the networking between individuals in an online environment (Chané, 2023). SNA systematically examines the connections between groups — or nodes — online. Through reverse engineering, SNA can determine the location and identity of each individual, making it clear that despite the use of pseudonymization to ensure anonymity, there remains a high chance that one would be unable to hide their own identity online.

Anonymization and pseudonymization reduce the risk of identifying the individual, but they do not guarantee absolute anonymity. The availability of information on social networking sites has led to an increase in de-anonymization attacks on sites such as LinkedIn or Xing (Wondracek et al., 2010). Similar to SNA, the information obtained from an individual's social groups on networking sites can enable complete identification of an individual, regardless of whether personal information is explicitly disclosed. Considering these issues, it is evident that anonymization and pseudonymization still have certain limitations. Therefore, exploring alternative solutions is imperative to achieve a balance between surveillance and privacy.

Establishing a Comprehensive Nationwide Privacy Law

Privacy protection in the US is complicated by the absence of a comprehensive federal privacy law. There is a patchwork of federal laws related to specific sectors and types of sensitive information, such as health and financial data, but these laws are not sufficiently broad or current (O'Connor, 2018). Although the Privacy Act of 1974 laid an important foundation for privacy protection, it is widely considered insufficient for addressing the complex privacy issues of the 21st century, indicating that updates or new legislation are needed to provide effective privacy safeguards. There are also state laws regulating privacy and they do not always align with one another, leading to conflicts and confusion among them (German Marshall Fund of the United States, 2021). To ensure that the U.S. can conduct surveillance in a transparent and privacy-respecting manner, the U.S. government should consider establishing a comprehensive federal privacy law modeled after the GDPR. Furthermore, American and international companies doing business in the EU are already required to comply with GDPR.

The GDPR, a stringent privacy and surveillance law, took effect in the EU on May 25, 2018 (GDPR.eu, n.d.). It consists of various articles concerning data protection, accountability, data security, and data processing. According to Article 5.1-2 of the GDPR on data protection, there are seven principles for lawful data processing: 1) processing must be legal and transparent to the data subject. the individual whose data is being captured; 2) data processing must be legitimate and the purposes for it must be defined; 3) only sufficient data should be processed — no excess data may be collected without reason; 4) the data must be accurate and up to date; 5) personal information may only be stored for the amount of time necessary for the specified purpose; 6) data processing should be encrypted to ensure privacy; and 7) the data processor should be held accountable to follow the rules of the GDPR with the certainty of being fined if they do not (GDPR.eu, n.d.). Although the aforementioned rules only represent some aspects of the GDPR's extensive articles, adopting just these seven rules in the U.S. would result in significant progress in terms of the public's trust in the government. Notably, imposing large fines for noncompliance would encourage the practices of government surveillance agencies to follow the privacy laws (O'Connor, 2018). Even though these government surveillance agencies — like the NSA — have substantial funding, the cost of surveillance would add up considering the vastness of the population being surveyed. These fines would also work especially well with small- or medium-sized agencies with more restricted monetary supply. Additionally, the GDPR

has demonstrated minimal interference with data processing for security purposes (Coyne, 2019). Considering all factors, establishing a comprehensive law similar to GDPR in the U.S. is a promising course of action to enhance privacy protection and bolster public confidence.

However, there are still issues with the GDPR. Although the GDPR permits the processing of some non-anonymized data, there is a lack of clarity on what boundaries should not be crossed (Siegert et al., 2020). Metadata on age, sex, and personality traits are extremely valuable for behavior analysis, but require non-anonymized data to be useful. Additionally, collecting data on interactions between people is also very valuable in contemporary times — a major reason why the NSA was performing secret data collection on U.S. citizens, which clashes with the GDPR. As the GDPR doesn't explicitly state the boundaries, it presents challenges for surveillance agencies attempting to collect data legally. Nevertheless, the GDPR has demonstrated many successes to this date and remains the most comprehensive solution available.

If the U.S. seeks to establish a law akin to the GDPR, it must explicitly define the legal boundaries regarding personal data collection. Echoing Franklin Delano Roosevelt's inaugural address, "an unprecedented demand and need for undelayed action may call for temporary departure from that normal balance of public procedure" (Roosevelt, 1933). In light of the pressing concerns of privacy and security in this day, we must similarly pivot from current surveillance policies in order to make progress in the protection of individual privacy.

Conclusion

The relationship between surveillance and security is complex and multi-faceted (Ünver, 2018). With the rapid advancements in technology, digital surveillance has become a pressing issue worldwide. Enforcing privacy in an era of pervasive digital surveillance presents a significant challenge that warrants comprehensive and thoughtful discussion. Edward Snowden's whistleblowing drew global attention to undisclosed mass surveillance, prompting the emergence of various privacy legislations and measures to protect individuals from the unwarranted covert surveillance. However, existing solutions, such as anonymization and pseudonymization, are insufficient as standalone strategies to ensure full privacy, given the possibility that additional information can be used to re-identify an individual.

By establishing a comprehensive nationwide privacy law similar to the GDPR framework, the U.S. could restore public trust and ensure that government surveillance is transparent and secure. While the GDPR has some flaws, particularly regarding the clarity of data processing, the U.S. has the opportunity to address the concerns by articulating precise provisions. A GDPR-like framework will be able to address the current absence of a comprehensive nationwide privacy law in the U. S.. That new regulation would be a significant step in upholding the right to privacy, a right that is foundational to true freedom as affirmed by the U. S. Constitution. Therefore, the establishment of a comprehensive privacy protection law will be of utmost importance for upholding this fundamental right and safeguarding the liberties of citizens in the United States.

Works Cited

- Art. 4 GDPR - Definitions. (2018, March 29). General Data Protection Regulation (GDPR). Retrieved February 25, 2024, from <https://gdpr-info.eu/art-4-gdpr/>
- Austin, L. M. (2015). Lawful Illegality: What Snowden Has taught Us about the Legal Infrastructure of the Surveillance State. In M. Geist (Ed.), *Law, Privacy and Surveillance in Canada in the Post-Snowden Era* (pp. 103–126). University of Ottawa Press. <https://doi.org/10.2307/j.ctt15nmj3c.8>
- Bauman, Z., Bigo, D., Esteves, P., Guild, E., Jabri, V., Lyon, D., & Walker, R. B. J. (2014). After snowden: Rethinking the impact of surveillance. *International Political Sociology*, 8(2), 121-144. <https://doi.org/10.1111/ips.12048>
- Chané, A.-L. (2023). Social Network Analysis. In F. Badache, L. R. Kimber, & L. Maertens (Eds.), *International Organizations and Research Methods: An Introduction* (pp. 248–254). University of Michigan Press. <http://www.jstor.org/stable/10.3998/mpub.11685289.54>
- Colvin, N. (2018). Whistle-Blowing as a form of digital resistance: State crimes and crimes against the state. *State Crime Journal*, 7(1). <https://doi.org/10.13169/statecrime.7.1.0024>
- Coyne, H. (2019). The Untold Story of Edward Snowden's Impact on the GDPR. *The Cyber Defense Review*, 4(2), 65–80. <https://www.jstor.org/stable/26843893>
- Data Protection Commission. (n.d.). Anonymisation and pseudonymisation: Data Protection Commission. Anonymisation and pseudonymisation | Data Protection Commission. <https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation>
- European Data Protection Supervisor. (n.d.). 10 Misunderstandings Related to Anonymisation. https://www.edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf.
- Feldstein, S. (2019). Distinguishing Between Legitimate and Unlawful Surveillance. In *The Global Expansion of AI Surveillance* (pp. 11–13). Carnegie Endowment for International Peace. <http://www.jstor.org/stable/resrep20995.6>
- GDPR.eu. (n.d.). What is GDPR, the EU's new data protection law? <https://gdpr.eu/what-is-gdpr/>.
- German Marshall Fund of the United States. (2021). *Rebuilding Trust in the Digital Ecosystem: New Mechanisms for Accountability: International Digital Accountability Council and GMF Digital*. German Marshall Fund of the United States. <http://www.jstor.org/stable/resrep30234>
- Greenwald, G. (2013). NSA collecting phone records of millions of Verizon customers daily. <https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order>
- Hagen, J., & Lysne, O. (2016). Protecting the digitized society—the challenge of balancing surveillance and privacy. *The Cyber Defense Review*, 1(1), 75–90. <http://www.jstor.org/stable/26267300>
- Kampmark, Binoy. (2014). Restraining the Surveillance State: A Global Right to Privacy. *Journal of Global Faultlines*, 2(1), 1–16. <https://doi.org/10.13169/jglobfaul.2.1.0001>

- NIST. (n.d.). Personal Data. personal data - Glossary | CSRC.
https://csrc.nist.gov/glossary/term/personal_data
- O'Connor, N. (2018). Reforming the U.S. Approach to Data Protection and Privacy. Council on Foreign Relations. <http://www.jstor.org/stable/resrep29972>
- Roosevelt, Franklin Delano, 1882-1945. Franklin D. Roosevelt's Inaugural Address of 1933. Washington, DC: National Archives and Records Administration, 1988.
- Siebert, I., Silber-Varod, V., Carmi, N., & Kamocki, P. (2020). Personal data protection and academia: GDPR issues and multi-modal data-collections "in the wild." *Online Journal of Applied Knowledge Management*, 8(1), 16-31.
[https://doi.org/10.36965/ojakm.2020.8\(1\)16-31](https://doi.org/10.36965/ojakm.2020.8(1)16-31)
- Thomas, T., & Chaleff, I. (2017). Moral Courage and Intelligent Disobedience. *InterAgency Journal*, 8(1), 58-66.
<https://thesimonscenter.org/wp-content/uploads/2017/03/IAJ-8-1-Winter2017-pg58-66.pdf>.
- Ünver, H. Akin. Politics of Digital Surveillance, National Security and Privacy. Centre for Economics and Foreign Policy Studies, 2018. JSTOR,
<http://www.jstor.org/stable/resrep17009>. Accessed 17 Aug. 2024.
- Wondracek, G., Holz, T., Kirda, E., & Kruegel, C. (2010). A Practical Attack to De-anonymize Social Network Users. 2010 IEEE Symposium on Security and Privacy, 2238-238.
<https://doi.org/10.1109/SP.2010.21>
- Wong, C. M., & HUMAN RIGHTS WATCH. (2015). Internet at a Crossroads: How Government Surveillance Threatens How We Communicate. In *World report 2015: Events of 2014* (1st ed., pp. 14–26). Bristol University Press. <https://doi.org/10.2307/j.ctvndv4w1.5>

Overview of Cerebrospinal Fluid, Glymphatic System, and Diagnosis of Multiple Sclerosis Through Machine Learning Extrapolation By Ethan Ngo

Abstract

The glymphatic system is a newly discovered process in the subarachnoid space that regulates metabolic clearance in the brain. It can potentially revolutionize the scientific understanding of sleep, cognitive function, and CNS disorders. The role of cerebrospinal fluid and its evolutionary origins will be investigated concerning the subarachnoid space. Moreover, this article will review and discuss the entirety of the glymphatic system, specifically the processes that occur in the perivascular space. Certain glial cells and proteins will be highlighted, such as astrocytes and AQP4, to further explain fluid transport and entry into the lymphatic network. The purpose of sleep and the accumulation of harmful CSF by-products will be closely scrutinized, particularly to examine glymphatic impairment and potential CNS disorders. Through CSF biomarker identification, current technology will help diagnose and identify neurodegenerative diseases. The new rise in artificial neural networks has innovated the possibilities of machine learning. Artificial intelligence algorithms, such as Random Forest, will extrapolate data on prevalent neurobiological factors of multiple sclerosis (MS) patients. The program will also demonstrate machine learning capabilities by predicting MS diagnosis in males and females. The new age of neurotechnology will feature more tangible and limitless roles in biomarker identification of CNS disorders. This article aims to explore the glymphatic system and cerebrospinal fluid with future technological innovations.

Section 1- Cerebrospinal Fluid and Glymphatic System

Introduction

In the central nervous system (CNS), the cerebrospinal fluid (CSF) circulates the subarachnoid space, which is located between the arachnoid mater and the pia mater. Composed primarily of blood plasma, it travels with the interstitial fluid (ISF) through hydrostatic and osmotic exchange. CSF contains albumin, immunoglobulins, and blood-derived proteins for metabolic recovery and brain protection. It is produced in various channels of the ventricular system and then released into the meninges; the layer of dura, pia, and arachnoid mater that encloses the brain (Di Terlizzi & Platt, 2006). In the subarachnoid space, CSF provides nutrients for neurons and is heavily involved in metabolic regulation. (Benveniste et al., 2017). As CSF travels throughout the brain, it accumulates harmful by-products that can lead to neurodegenerative diseases. The newly discovered CNS ‘waste clearance’ system, the glymphatic system, has been recently associated with several disorders, one of which is Multiple Sclerosis (MS). Abnormalities and impairment in glymphatic function have led to MS demyelination and acute symptoms in progressive stages of clinical disability. Defects in the glymphatic system mediate underlying pathological mechanisms in multiple sclerosis and other neurodegenerative ailments (Carotenuto et al., 2022). The modern innovation of machine

learning in healthcare and data analysis of CSF biomarkers can aid with the diagnosis of CNS disorders and help procure solutions.

Evolutionary Origins of CSF in Ancestral Lineage

CSF descends from a deuterostome lineage of the ancestral organisms Abulacracia and Chordata. The original function of CSF was to contribute to a balanced chemical environment for nonsynaptic signal communication. However, as evolution occurred, dendritic processes were formed in CSF brain ventricles and led to the formation of terminals (stereocilia). These passageways allowed for CSF-contacting neurons to enter synaptic zones in the brain. The development of CSF influx further continued after the massive clade formed between the deuterostome ancestors (Bueno & Garcia-Fernández, 2016).

The Abulacracia family's descent into invertebrates resulted in the species' limited cognitive capabilities due to the absence of a central nervous system. Meanwhile, the Chordata organism evolved into Urochadata, Cephalachorta, and Vertebrata—the Vertebrata species, the last of the evolutionary line, includes fish, amphibians, reptiles, birds, and mammals. In other words, the Vertebrata species, an organism with a backbone, contains CSF to help protect its brain and spinal cord.

The Vertebrata species have neuronal communication and provide nutrients to cells through CSF barrier transfer systems. The neuronal expansion led to the distribution of CSF functions and the creation of emergent properties. Due to the development of neuron complexity, CNS neurons do not directly detect environmental conditions; however, they synapse with other neurons in sensory organs. Overall, the evolution of CSF created layers of neurons and the formation of an advanced neocortex.

Ventricular System: Synthesis and Circulation of CSF

The ventricular system, located throughout the brain parenchyma, is a series of connected cavities that synthesizes and secretes cerebrospinal fluid to the subarachnoid space and the spinal cord (Shenoy & Lui, 2023). The subarachnoid space lies between two layers of the meninges: the pia mater and the arachnoid mater. CSF is produced by a densely folded vascularized structure known as the choroid plexus. The choroid plexus is a group of epithelial cells that lie on the basolateral membrane and connect to surrounding vessel walls. It is facilitated through ion transport channels which utilize an osmotic gradient to provide water and nutrients from fenestrated blood capillaries. At the apical membrane, motile cilia help move CSF through the ventricular system (Javed et al., 2023).

Once CSF is produced, it travels from the lateral ventricle to the third ventricle by the interventricular foramen. The fluid then enters the cerebral aqueduct and flows into the fourth ventricle. Since each ventricle contains a choroid plexus, a convective flow leads to a pressurized pathway. At this point, CSF can exit from four passageways: right aperture, median aperture, left aperture, and central canal brain stem. Finally, CSF circulates the subarachnoid space where it provides nutrients and protection for the brain (Khasawneh et al., 2018).

Function and Purpose of CSF

I. Protection of the brain and spinal cord

During head and neck injuries, CSF acts as a shock absorber between the brain and skull, thus preventing increased intracranial pressure (Telano & Baker, 2023). Essentially, CSF ensures minimal damage to the functional tissue and maintains an individual's cognitive ability.

II. Nutrients for brain parenchyma

CSF provides nutrients to neuron cells and glial cells. These nutrients include glucose, protein, and minerals. Moreover, CSF helps provide an ionically balanced environment to promote normal neuron-synaptic exchange.

III. Metabolic Waste Removal

CSF utilizes the glymphatic system to irrigate unhealthy by-products in CNS (Veening & Barendregt, 2010). Sinuses are another route for CSF to enter the lymphatic system and be refreshed.

The Glymphatic System

The glymphatic system is the brain's metabolic waste disposal. It utilizes the convective flow of CSF and interstitial fluid (ISF), also known as the glymphatic influx, to clear waste products and reduce the amount of toxic by-products. The glymphatic system utilizes capillaries, astrocytes, and sinuses to irrigate CSF (Benveniste et al., 2017). If not regulated correctly, by-products can accumulate within the CSF and create fatal repercussions. Malfunctions in the glymphatic system can contribute to several neurodegenerative diseases.

CSF Process in Glymphatic System (Perivascular Space)

At the cortical surface of the brain, CSF travels along the pial artery. The fluid is facilitated by smooth muscle cells and an influx gradient. The pial artery then penetrates the brain parenchyma and enters the perivascular space, otherwise known as the Virchow-Robin space (Figure 1). In this periarterial region, the arterial pulsations help drive CSF down the basal lamina, a sheet of extracellular matrix (Jessen et al., 2015).

The basal lamina continues to enter the parenchyma where it eventually diminishes into a capillary region. At the base of the basal lamina, a tight junction binds the vascular endfeet of astrocytes (Figure 1). Astrocytes are glial cells in the CNS that promote metabolic and neuroprotective functions.

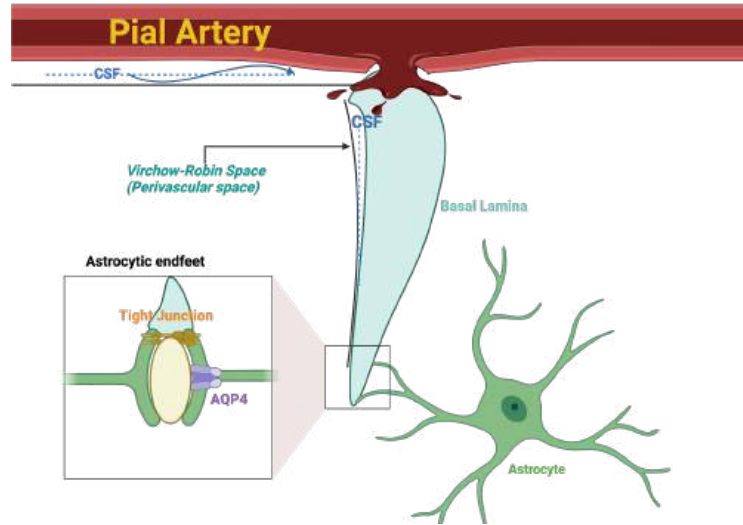


Figure 1: Glymphatic Process (Perivascular Space)

In the astrocytic endfeet, a protein channel, aquaporin-4, facilitates the interstitial (ISF) and CSF exchange (Mestre et al., 2020). AQP4 then excretes the CSF-ISF fluid into the neuronal tissue and generates a glymphatic influx. The overall pressure gradient of CSF allows for a convective flow of ISF into the perivenous space, a fluid-filled structure traveling along veins in the perivascular space (Figure 2). In this region, interstitial fluid enters the cervical lymph system to cleanse off harmful by-products (Jessen et al., 2015).

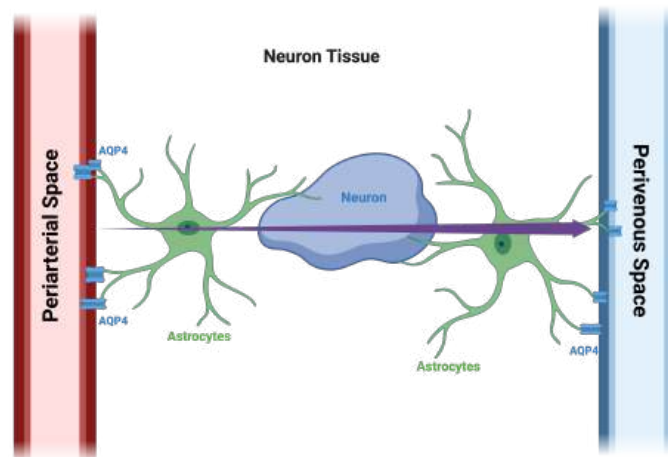


Figure 2: Astrocytic Polarization in Periarterial to Perivenous Space

Overall, the glymphatic system is a neurovascular process involving constant CSF-ISF exchange. Glymphatic activity is heavily reliant on astrocytes and AQP4. These glial bodies are the main facilitators of glymphatic influx and convective fluid transportation.

Role AQP4

Aquaporin-4 channels are proteins that constantly exchange interstitial and cerebrospinal fluid. Expressed by glial cells, AQP4 plays an essential role in draining CSF-ISF fluid into

perivascular pathways (Gao et al., 2023). These protein channels produce a polarity and flow into the neuron tissue; CSF is exchanged across the subarachnoid space (Jessen et al., 2015).

AQP4 often decreases in functionality as an individual ages. For instance, AQP4 utilizes osmotic pressure to produce a current. However, these aquaporin channels degrade over time. The polarization and concentration significantly decrease in its dispersal of influx gradients. Once this occurs, CSF-ISF exchange becomes sporadic, and the glymphatic system becomes impaired (Jessen et al., 2015). The disruption of CSF-ISF fluid transportation from the periarterial to the perivenous space greatly increases B-amyloid protein and other harmful products (Iliff & Simon, 2019). Without clear irrigation from the lymphatic system, neurological diseases like Alzheimer's tend to rise more frequently.

Sleep

The glymphatic system is optimal during sleep. The system constantly detoxifies the brain yet undergoes certain cycles of efficient production. In the daytime, it remains generally inactive. However, during certain stages of sleep, like REM, chemicals involved in homeostasis help drive CSF irrigation throughout the brain.

The glymphatic system is also regulated through a chief modulator known as norepinephrine. During the daytime, this chemical is constantly released into the bloodstream to activate cognitive function, attention, and arousal (Hussain et al., 2023). Thus, during certain stages of sleep, the central acuity of norepinephrine dramatically decreases. Arousal levels minimize and expand the extracellular and interstitial space. Once this occurs, interstitial fluid and CSF influx increase because of less resistance from surrounding walls (Reddy & Van Der Werf, 2020). Greater volume of ISF and CSF leads to greater metabolic waste disposal and a rise in glymphatic function during sleep.

Moreover, the glymphatic system operates in tandem with respiratory signals and intracranial oscillations. Low productivity in the respiratory system causes stable blood pressure levels and a consistent heart rate. In the most efficient stage of the glymphatic system, N3, amyloid beta removal is optimal, arterial pulsations are minimal, and the drive for CSF is relatively high (Reddy & Van Der Werf, 2020). At this stage, copious amounts of amyloid beta are removed and detoxified. In general, this process is extremely delicate; minimal changes to arterial pulsations can decrease glymphatic influx and toxin disposal.

Sinuses: Lymphatic meningeal vessels

Moreover, the glymphatic system has an alternative route for CSF irrigation. In the meningeal layers, there are several air-filled cavities that are known as sinuses. The more commonly used ones are the superior sagittal sinus and the transverse sagittal sinus. These sinuses contain arachnoid granulations that serve as a pathway into lymph meningeal vessels of lymphatic systems (Letchuman & Donohoe, 2023). The CSF is then rinsed throughout the brain and exits via the sinuses; the lymphatic systems replenish and cleanse the CSF.

Impairment to Glymphatic System (CSF Biomarkers)

As CSF travels throughout the ventricular system, it collects unwanted by-products that can severely damage neuron cells. If not treated carefully, these products can cause impediments in motion capabilities and mental acuteness (Carotenuto et al., 2022).

In order to diagnose neurodegeneration, scientists generally use CSF biomarkers to determine if the disorder is present in the patients (Mattsson-Carlgren et al., 2022). There are three main CSF biomarkers:

- I. B-amyloid
- II. Tau Protein
- III. $A\beta_{42}$

Common Technologies Used in CSF Biomarker Tracing

I. Mass Spectrometry

Innovative technology that is used to determine a type of mass. Commonly used in proteomics to identify proteins and classify chemical structures. For mass spectrometry to function, the source or protein must be ionized (a substance that contains charged particles). With an analyzer, molecules are split from the mass. Scientists can then detect these ionized atoms in the protein (Parker et al., 2010). Lastly, the device amplifies the result and retrieves quantifiable signals from the charged particle.

Mass spectrometry also utilizes two different performance outputs: unbiased vs. targeted. Targeted mass spectrometry is high-throughput and sensitive. It uses a process known as parallel reaction monitoring (Bharucha et al., 2019). On the contrary, an unbiased approach is the proteomic study of novel biomarkers and large-scale proteins. Through this holistic review, study cohorts are created based on distinction and expression divergence.

II. ELISA

Enzyme-Linked Immunosorbent Assay, ELISA, is an analytics biochemical assay that utilizes antibodies and ligand signals to detect the immunological response of a particular antigen. ELISA binds the antibody to the foreign object and monitors the change in gene expression. In general, technology plays a major role in the diagnosis of immunological-compromised diseases.

For instance, ELISA can diagnose Parkinson's disease. Tyrosine hydroxylase, an oxygenase, is heavily associated with the pathogenesis of various disorders in the brain. Through a targeted approach, the assay can monitor TH expression with neuromodulatory agents; these antibodies can provide measurable analysis of any developmental changes in the body (Fauss et al., 2013). Other diseases that ELISA can detect include, but are not limited to, HIV and neurocysticercosis (Rosas et al., 1986).

III. Two-photon microscopy

Two-photon microscopy is an imaging technique that utilizes fluorescent lighting to trace chemical biomarkers. First, tracers are injected into the *vivo* sections of the brain parenchyma to create a depictive graph. Next, an invasive procedure is required to perform two-photon microscopy. This includes the opening of a cranial window and skinning of the skull. However, the amount of exposure and field of view is extremely narrow. Once the confocal microscopy is placed in the aperture, two photons are fired into the targeted area to visualize the fluorescent tracers on biomarkers.

CSF Neurodegenerative Diseases

CNS neurodegenerative diseases are heavily influenced by impairments of CSF flow and the glymphatic system (Simon & Iliff, 2016). The five most common disorders caused by CSF-related factors include:

- ❖ Alzheimer's
- ❖ Multiple Sclerosis (MS)
- ❖ Hydrocephalus
- ❖ Parkinson Disease
- ❖ Amyotrophic Lateral Sclerosis (ALS)

Section 2: Machine Learning Extrapolation of Biomarkers in Multiple Sclerosis

Introduction of Machine Learning

Machine learning is the use and development of computer systems. New technology has allowed for machine learning to analyze data, draw inferences, and identify trends. Essentially, machine learning adapts by continuously learning test sets and eliminates the need for explicit instructions (Baştanlar & Özuysal, 2014).

Role of Machine Learning and Artificial Intelligence

The development of deep learning (DL) has expanded the medical limitations that once bounded artificial intelligence. DL has opened a new gate of complex technologies and innovation of practical uses. Before, algorithms possessed only three to five layers of neural networks. With this new type of machine learning, AI has expanded to ten layers. Now, there are millions more artificial neurons and AI capacities (Bohr & Memarzadeh, 2020). With this rise, the application of machine learning in healthcare is undeniable.

AI has recently become popular in administrative offices, robotic surgeries, and clinical image analysis. Precision medicine, such as omics-based tests, has used machine learning programs to predict treatments and generate correlations from population pools. The process of drug discovery and molecular target identification, like SMILES³, has been simplified (Bohr & Memarzadeh, 2020). At Johns Hopkins University, smart tissue autonomous robots, otherwise known as STAR, have outperformed surgical doctors in bowel anastomosis procedures (Graham, 2022).

³ simplified molecular input line entry system

Moreover, deep learning and medical scanning recognition allow for accurate data collection. Influenced by DL techniques, convolutional neural networks (CNNs) have provided a new interpretation of radiological images in the human visual cortex. In general, AI can be a flexible structure for healthcare systems.

Introduction to Multiple Sclerosis

Multiple sclerosis, otherwise known as MS, is an autoimmune inflammatory disease that attacks neuronal tissue and leads to cognitive dysfunction. The immune system targets the myelin sheath on the axon of the neuron. Neuronal communication becomes impaired and causes a cascade through the brain stem (Barkhane et al., 2022).

MS can be caused by a combination of environmental factors and genetic inheritance. In recent studies, individuals who have relatives with MS are more susceptible to the disorder (NINDS, 2023). Females have a higher chance of MS, 3:1 ratio, than men due to their abrupt hormonal imbalances during puberty, pregnancy, puerperium, and menopause. However, testosterone in men leads to more serious cases of this condition (Ysrraelit & Correale, 2019). Additionally, MS tends to target people with a lack of exposure to sunlight or those lacking vitamin D (NINDS, 2023).

MS is the result of “foreign” objects being misidentified in the body. If the blood-brain barrier is damaged, white-blood cells will flood into the brain stem. Thus, the immune system will misinterpret myelin sheaths as “foreign” and utilize T-cells to break down these structures (Miron, 2019).

There are four main types of MS:

- I. Relapse-remitting MS
 - A. Periods of recovery and activity of symptoms
 - B. The duration of the attack varies
 - C. The most common form of the disease
- II. Secondary-progressive MS
 - A. Relapse of MS attacks
 - B. Gradual degradation of symptoms
- III. Primary-progressive MS
 - A. Progressive symptoms
 - B. No form of relapse
- IV. Progressive-relapsing MS
 - A. Acute relapses
 - B. Progressive symptoms
 - C. The rarest form of the disease

Extensive Biological Overview of MS

Oligodendrocytes are organisms that synthesize myelin and protect the axonal pathway of a neuron. In MS, rogue T-cells, a form of lymphocytes, first bind onto the myelin sheath. This

binding causes the release of cytokines or small signaling proteins such as IL-1, IL-6, TNF- ∞ , and INF- γ . These cytokines facilitate the dilation of the blood vessels (Figure 3).

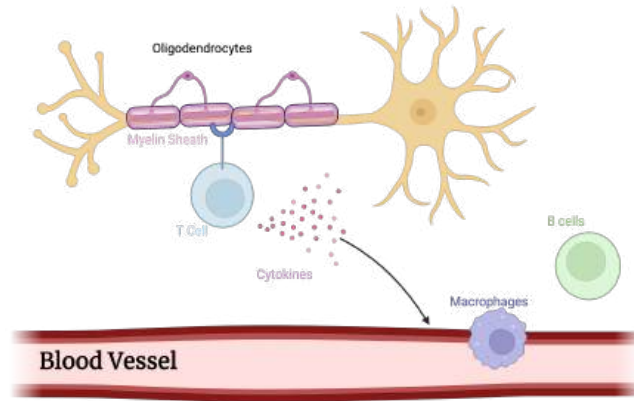


Figure 3: Stage I of Myelin Deterioration in MS

Once the permeable barrier between the blood vessel and the brain dilutes, an inflammatory response will occur. This will allow for macrophages and B-cells to enter the brain. Macrophages are digestive proteins that engulf microorganisms to clean out dead cells and toxic products. B-cells are lymphocytes that help facilitate this action by producing antibodies. These antibodies bind to the microorganism and neutralize it. Macrophages will use these antibodies as a signal and detect the defective cell. In MS, the cell is not “defective” but rather caused by mis-stimulation of rogue T cells.

Macrophages will then engulf the oligodendrocytes; demyelination of the neuron will occur. Once the sheath decays and the axon becomes exposed, neuron signals can be easily interfered with. The disintegrated sheaths will form plaques known as sclera (Figure 4). Thus, cognitive function and motile ability become impaired through the progressive decay of neurons’ signal transduction.

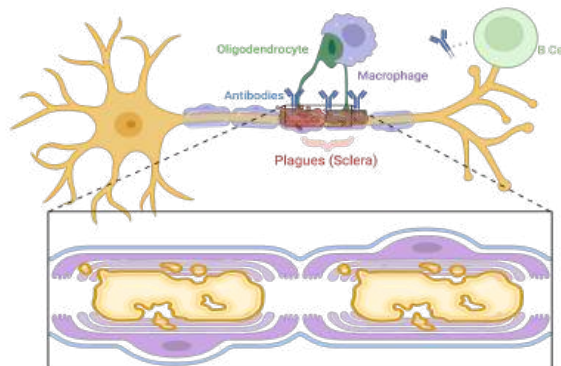


Figure 4: Stage II of Myelin Deterioration in MS

Oligoclonal Bands

Oligoclonal bands (OGBs) are immunoglobulin proteins that surface in the CSF due to an inflammatory response. These bands are key detectors in CNS diseases. In MS, when T-cells begin to attach to the myelin sheath, they send cytokines (inflammatory markers) that help B-cells pass through the blood-brain membrane. Oligoclonal bands are then generated and sent to the CNS (Neilson et al., 2022).

However, oligoclonal bands (OGBs) are also present in different neurological diseases. OGBs are present in multiple sclerosis, herpes simplex encephalitis, and subacute sclerosing panencephalitis. In some exceptions, there are positive bands in Alzheimer's, amyotrophic lateral sclerosis, and polyneuropathy (Chu et al., 1983). Essentially, OGBs help diagnose a CNS immunologically compromised disease but do not indicate the specific condition.

Introduction to Dataset and Machine Learning Proposal

In a cohort case study conducted on Mexican mestizo patients, scientists analyzed the presence of CSF oligoclonal bands, as well as auditory, sensory, and motor reactions to various simulations. These indicators were collected in a shotgun assay, providing an accurate diagnosis of CNS disorder: multiple sclerosis. Two test groups were held to record data on patients' backgrounds and physical exams (Benjamin Pineda, 2023).

The first group, CDMS, were clinically diagnosed patients of MS and documented their score in the Expanded Disability Status Score (EDSS)⁴. The second group, non-CDMS, were non-diagnosed patients of MS. Their symptoms were recorded; however, no values or documentation of their EDSS score were available.

In the project, machine learning will identify and diagnose non-CDMS patients. By using artificial intelligence, a trainable and testable algorithm will read the code and provide an accurate prognosis. Furthermore, the AI will identify neurobiological factors and their importance to the diagnosis of MS in both female and male patients. The consensus involves machine learning, its application and purpose in healthcare, and the demonstration of the new innovative age of biological artificial intelligence.

Data Pre-Processing

To first analyze the dataset, several packages and Python libraries were imported. For example, numpy and pandas were used to store numerical information and allow modification of the data to fit in an array. The public dataset collected by the cohort study was uploaded and stored in a pandas data frame. This is a standard process that ensures advanced mathematical computations in a relational way. Moreover, seaborn and matplotlib.pyplot were used to display the data visually throughout the code. Matplot was primarily utilized in illustrating the variances of MS relative importance.

Before the program was able to run properly, several data pre-processing steps were required. Several columns were dropped: initial EDSS, age, breastfeeding, and schooling. These

⁴ a quantifiable scale on monitoring the disabling effects of MS in the patient

columns were used for background information on the patient and had no relevant importance in the diagnosis of MS. To predict non-CDMS patients accurately, the program had to solely focus on neurobiological components.

Due to multiple sclerosis's changes in behavior in different sexes, the dataset was copied and segmented into two parts. This action was necessary since the biological relative importance is scaled differently in males compared to factors in females. Moreover, by splitting the code, the predicted EDSS can be more precisely calculated since males tend to have a higher score on the disability scale.

Relative Importance Interpretation

Using scikit-learn, a machine learning tool for predictive data analysis, the female data set was split into a train and test set. The StandardScaler() method was then imported and implemented into the code to restructure the dataset's features. The deep learning algorithm, Random Forest, created a decision tree based on the test and train set. This allowed it to determine trends in certain factors that have the most effect on the diagnosis of MS.

I. Females

Interestingly, machine learning revealed that the most important factor in identifying MS for females was the periventricular MRI (Figure 5). This MRI locates periventricular white matter, a network of nerve fibers, in the frontal, parietal, temporal, or occipital part of the brain.

Studies demonstrate that the periventricular white matter in the cerebellum controls sensory information. MS demyelination in the cerebellum can cause lesions, abnormal or damaged areas, in the spinal cord. This emphasizes destabilizing motor functions in the brainstem. Therefore, women have common MS symptoms of dizziness and loss of balance (Filippi et al., 2019). Periventricular white matter lesions can also lead to higher chances of neurodegenerative diseases like dementia and Alzheimer's. Most patients with MS also experience pseudobulbar palsy or difficulties in upper motor movement (Casini, 2013).

The initial symptoms have less effect on MS diagnosis than males. This is because women generally have a relapse-remitting MS, a mild form of the disorder.

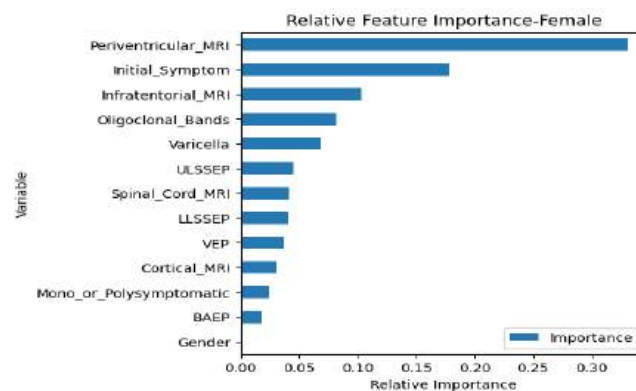


Figure 5: Relative Feature Importance in Females

II. Males

The same process of splitting the dataset and training it through Random Forest was utilized for the male test case. In comparison to the females, initial symptoms and the infratentorial MRI had more significance in the diagnosis of MS (Figure 6).

The infratentorial white matter located in the cerebellum facilitates muscle control, movement, tremor, memory, and balance. Located in the brain stem, the infratentorial white matter affects language processing and motor instability. As lesions in the spinal cord arise, individuals experience a progressive disability. This could be a possible justification for men having greater neurodegeneration and primary-progressive MS (Dekker et al., 2020).

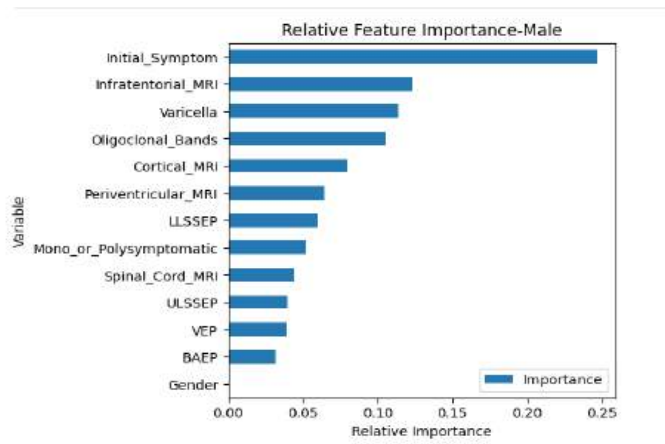


Figure 6: Relative Feature Importance in Males

Although men have a lower rate of acquiring multiple sclerosis, the symptoms that they exhibit are much more acute than women. This correlates to machine learning's prediction of initial symptoms playing a vital role in a patient's score on the EDSS scale.

Clearly, machine learning is efficient in identifying the biological factors of CNS disorders. Artificial intelligence was able to track trends and use a training set to calculate varying MS indicators in different sexes. It was able to reveal key qualities, such as the periventricular and intraventricular white matter, that were not evidently noticeable.

Machine Learning Discoveries

As showcased in the program, machine learning abstraction can lead to multiple discoveries. For example, many studies indicated that the largest MS indicator was oligoclonal bands. Although this biological feature ranked high in both females and males, it was not the primary biomarker. Despite CSF being hypothesized as the major contributor to MS, the data implies white matter lesions play a much more significant role.

Moreover, artificial intelligence detected another component of multiple sclerosis: varicella. Varicella, also commonly known as chickenpox, has a "causal effect" on MS. Current research is investigating this unknown connection to find ways to prevent MS progression (Zhu et al., 2023).

Diagnosis of Machine Learning in Patients

Artificial intelligence can also predict the prognosis of a patient's condition if given a trained dataset. Random forest's built-in method, `forest.predict()`, using the test set as a parameter to predict MS in both CDMS and non-CDMS patients. The algorithm had an 85% accuracy in correctly identifying multiple sclerosis in non-CDMS individuals (Figure 7). With larger datasets and more advanced deep learning algorithms, predictions can be more precise. The results are displayed down below:

	precision	recall	f1-score	support
CDMS	0.62	0.79	0.70	19
Non-CDMS	0.85	0.72	0.78	32
accuracy			0.75	51
macro avg	0.74	0.75	0.74	51
weighted avg	0.77	0.75	0.75	51

Figure 7: Random Forest Prediction on MS Diagnosis

Conclusion: Benefits of Machine Learning in Healthcare

Machine learning has multiple positive applications for health care. In the diagnosis of serious health conditions, it eliminates human error like subjective testing. In most cases, scientists have bias over possible conditions based on initial symptoms. Trends demonstrate that humans naturally gravitate toward certain prognoses based on recency and their accustomed knowledge of the subject. With machine learning, data is empirically analyzed, and diagnosis is a direct reflection of these results (Habeheh & Gohel, 2021). Moreover, machine learning can dramatically improve the deliverability of solutions among a variety of patients. Not only is data analysis more accurate, but also more scalable; it reaches wider demographics in different communities. With more accessible solutions, resources can be expended carefully and efficiently (Davenport & Kalakota, 2019). The prescriptive purposes of machine learning foreshadow the new modern era of technology. Algorithms can be trained to plan alternative cognitive behavioral therapies; one that may be unconventional and personalized toward the patient. Health care will be revolutionized by the new world of biotechnology.

Proposal

Future innovations in artificial intelligence and machine learning will encourage scientific breakthroughs in cerebrospinal fluid and glymphatic systems. This newly-discovered process can further discoveries in neurodegenerative linkages and neurobiological metabolic recovery. With deep learning algorithms, biomarker investigation of cerebrospinal fluid can reveal more information on its roles and functions in the brain. By exploring CSF in the glymphatic structure, scientists will pioneer research on neuronal cognition and CNS diseases.

The establishment of neurotechnology will promote scalable solutions and manifest in a modern blueprint for healthcare systems.

Methods Section

To view the full code, please visit [EthanNgo_MSMLearning.ipynb](#) in GitHub public repository

Figure Legend

Fig. 1. Glymphatic Process (Perivascular Space)

The picture demonstrates the processes that are involved with the glymphatic system. CSF travels from the perivenous space, through the basal lamina, and into the astrocytic endfeet. The tight junction facilitates the CSF transportation from the extracellular sheet into the subarachnoid space. Then, CSF and ISF are pumped into the neuron tissue, eventually leading down to the perivenous space. *Created with BioRender.com*

Fig. 2. Astrocytic Polarization in Periarterial to Perivenous Space

AQP4 channels are gated at the astrocytic endfeet. The astrocytes attach to both the periarterial and the perivenous walls. When CSF is pumped into the subarachnoid space and neuron tissue, the fluid follows a pressure gradient produced by these water channels. However, as aging occurs, this pressure gradient degrades because of the deterioration of AQP4. *Created with BioRender.com*

Fig. 3. Stage I of Myelin Deterioration in MS

T-cells first bind onto the myelin sheath of the oligodendrocytes. Cytokine chemicals then enter the bloodstream to signal macrophages and B-cells emerge. This is the first stage in the breakdown of myelin. *Created with BioRender.com*

Fig. 4. Stage II of Myelin Deterioration in MS

B-cells send antibodies onto the myelin sheath. This alerts the macrophages to attach to the oligodendrocytes. Gradually, the myelin sheath decays and forms into plaques, also known as the sclera. *Created with BioRender.com*

Fig. 5. Relative Feature Importance in Females

Random forest predicts the important factors that contribute to MS diagnosis in females. The most notable contributor was periventricular MRI which highlighted the role of periventricular lesions in MS symptoms.

Fig. 6. Relative Feature Importance in Males

Random forest predicts the important factors that contribute to MS diagnosis in males. The most notable contributors were initial symptoms and infratentorial MRI. In other words, severe neurodegeneration is caused by infratentorial lesions.

Fig. 7. Random Forest Prediction on MS Diagnosis

Deep learning correctly predicted MS diagnosis in non-CDMS patients with an 85% proficiency. However, in CDMS patients, random forest was predicted with 62% proficiency. A greater volume of data and advanced learning models will procure a more accurate prognosis.

Works Cited

- Barkhane, Z., Elmadi, J., Satish Kumar, L., Pugalenth, L. S., Ahmad, M., & Reddy, S. (2022). Multiple Sclerosis and Autoimmunity: A Veiled Relationship. *Cureus*.
<https://doi.org/10.7759/cureus.24294>
- Baştanlar, Y., & Özuysal, M. (2014). Introduction to Machine Learning. In M. Yousef & J. Allmer (Eds.), *miRNomics: MicroRNA Biology and Computational Analysis* (Vol. 1107, pp. 105–128). Humana Press. https://doi.org/10.1007/978-1-62703-748-8_7
- Benjamin Pineda. (2023). Conversion predictors of Clinically Isolated Syndrome to Multiple Sclerosis in Mexican patients: A prospective study. [dataset]. Mendeley.
<https://doi.org/10.17632/8WK5HJX7X2.1>
- Benveniste, H., Lee, H., & Volkow, N. D. (2017). The Glymphatic Pathway: Waste Removal from the CNS via Cerebrospinal Fluid Transport. *The Neuroscientist*, 23(5), 454–465.
<https://doi.org/10.1177/1073858417691030>
- Bharucha, T., Gangadharan, B., Kumar, A., De Lamballerie, X., Newton, P. N., Winterberg, M., Dubot-Pérès, A., & Zitzmann, N. (2019). Mass spectrometry-based proteomic techniques to identify cerebrospinal fluid biomarkers for diagnosing suspected central nervous system infections. A systematic review. *Journal of Infection*, 79(5), 407–418.
<https://doi.org/10.1016/j.jinf.2019.08.005>
- Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare* (pp. 25–60). Elsevier.
<https://doi.org/10.1016/B978-0-12-818438-7.00002-2>
- Bueno, D., & Garcia-Fernández, J. (2016). Evolutionary development of embryonic cerebrospinal fluid composition and regulation: An open research field with implications for brain development and function. *Fluids and Barriers of the CNS*, 13, 5.
<https://doi.org/10.1186/s12987-016-0029-y>
- Carotenuto, A., Cacciaguerra, L., Pagani, E., Preziosa, P., Filippi, M., & Rocca, M. A. (2022). Glymphatic system impairment in multiple sclerosis: Relation with brain damage and disability. *Brain*, 145(8), 2785–2795. <https://doi.org/10.1093/brain/awab454>
- Casini, G. (2013). Are Periventricular Lesions Specific for Multiple Sclerosis? *Journal of Neurology & Neurophysiology*, 04(02). <https://doi.org/10.4172/2155-9562.1000150>
- Chu, A. B., Sever, J. L., Madden, D. L., Iivanainen, M., Leon, M., Wallen, W., Brooks, B. R., Lee, Y. J., & Houff, S. (1983). Oligoclonal IgG bands in cerebrospinal fluid in various neurological diseases. *Annals of Neurology*, 13(4), 434–439.
<https://doi.org/10.1002/ana.410130410>
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98. <https://doi.org/10.7861/futurehosp.6-2-94>
- Dekker, I., Sombekke, M. H., Balk, L. J., Moraal, B., Geurts, J. J., Barkhof, F., Uitdehaag, B. M., Killestein, J., & Wattjes, M. P. (2020). Infratentorial and spinal cord lesions: Cumulative predictors of long-term disability? *Multiple Sclerosis Journal*, 26(11), 1381–1391.
<https://doi.org/10.1177/1352458519864933>

- Di Terlizzi, R., & Platt, S. (2006). The function, composition and analysis of cerebrospinal fluid in companion animals: Part I – Function and composition. *The Veterinary Journal*, 172(3), 422–431. <https://doi.org/10.1016/j.tvjl.2005.07.021>
- Fauss, D., Motter, R., Dofiles, L., Rodrigues, M. A. V., You, M., Diep, L., Yang, Y., Seto, P., Tanaka, K., Baker, J., & Bergeron, M. (2013). Development of an enzyme-linked immunosorbent assay (ELISA) to measure the level of tyrosine hydroxylase protein in brain tissue from Parkinson’s disease models. *Journal of Neuroscience Methods*, 215(2), 245–257. <https://doi.org/10.1016/j.jneumeth.2013.03.012>
- Filippi, M., Preziosa, P., Banwell, B. L., Barkhof, F., Ciccarelli, O., De Stefano, N., Geurts, J. J. G., Paul, F., Reich, D. S., Toosy, A. T., Traboulsee, A., Wattjes, M. P., Yousry, T. A., Gass, A., Lubetzki, C., Weinshenker, B. G., & Rocca, M. A. (2019). Assessment of lesions on magnetic resonance imaging in multiple sclerosis: Practical guidelines. *Brain*, 142(7), 1858–1875. <https://doi.org/10.1093/brain/awz144>
- Gao, Y., Liu, K., & Zhu, J. (2023). Glymphatic system: An emerging therapeutic approach for neurological disorders. *Frontiers in Molecular Neuroscience*, 16, 1138769. <https://doi.org/10.3389/fnmol.2023.1138769>
- Graham, C. (2022). Robot Performs First Laparoscopic Surgery Without Human Help. Johns Hopkins University: The Hub. <https://hub.jhu.edu/2022/01/26/star-robot-performs-intestinal-surgery/>
- Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291–300. <https://doi.org/10.2174/1389202922666210705124359>
- Hussain, L. S., Reddy, V., & Maani, C. V. (2023). Physiology, Noradrenergic Synapse. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK540977/>
- Iloff, J., & Simon, M. (2019). CrossTalk proposal: The glymphatic system supports convective exchange of cerebrospinal fluid and brain interstitial fluid that is mediated by perivascular aquaporin-4. *The Journal of Physiology*, 597(17), 4417–4419. <https://doi.org/10.1113/JP277635>
- Javed, K., Reddy, V., & Lui, F. (2023). Neuroanatomy, Choroid Plexus. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK538156/>
- Jessen, N. A., Munk, A. S. F., Lundgaard, I., & Nedergaard, M. (2015). The Glymphatic System: A Beginner’s Guide. *Neurochemical Research*, 40(12), 2583–2599. <https://doi.org/10.1007/s11064-015-1581-6>
- Khasawneh, A. H., Garling, R. J., & Harris, C. A. (2018). Cerebrospinal fluid circulation: What do we know and how do we know it? *Brain Circulation*, 4(1), 14–18. https://doi.org/10.4103/bc.bc_3_18
- Letchuman, V., & Donohoe, C. (2023). Neuroanatomy, Superior Sagittal Sinus. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK546615/>
- Mattsson-Carlgrén, N., Grinberg, L. T., Boxer, A., Ossenkuppele, R., Jonsson, M., Seeley, W., Ehrenberg, A., Spina, S., Janelidze, S., Rojas-Martinex, J., Rosen, H., La Joie, R., Lesman-Segev, O., Iaccarino, L., Kollmorgen, G., Ljubenkova, P., Eichenlaub, U., Gorno-Tempini, M. L., Miller, B., ... Rabinovici, G. D. (2022). Cerebrospinal Fluid

- Biomarkers in Autopsy-Confirmed Alzheimer Disease and Frontotemporal Lobar Degeneration. *Neurology*, 98(11), e1137–e1150. <https://doi.org/10.1212/WNL.0000000000200040>
- Mestre, H., Mori, Y., & Nedergaard, M. (2020). The Brain's Glymphatic System: Current Controversies. *Trends in Neurosciences*, 43(7), 458–466. <https://doi.org/10.1016/j.tins.2020.04.003>
- Miron, V. (2019). Immune system and MS. Multiple Sclerosis Society UK. <https://www.mssociety.org.uk/research/explore-our-research/emerging-research-and-treatments/immunomodulation#:~:text=In%20MS%2C%20the%20immune%20system,properly%20and%20may%20eventually%20die>
- Neilson, L. E., Hollen, C., Hiller, A., & Wooliscroft, L. (2022). Oligoclonal Bands in Multiple System Atrophy: Case Report and Proposed Mechanisms of Immunogenicity. *Frontiers in Neuroscience*, 16, 852939. <https://doi.org/10.3389/fnins.2022.852939>
- NINDS. (2023). National Institute of Neurological Disorders and Stroke: Multiple Sclerosis. U.S. Department of Health and Human Services. <https://www.ninds.nih.gov/health-information/disorders/multiple-sclerosis#:~:text=Research%20shows%20that%20genetic%20vulnerabilities%20combined%20with%20environmental%20factors%20may%20cause%20MS.&text=MS%20itself%20is%20not%20inherited,relative%20who%20also%20have%20MS.>
- Parker, C. E., Warren, M. R., & Mocanu, V. (2010). Mass Spectrometry for Proteomics. In O. Alzate (Ed.), *Neuroproteomics*. CRC Press/Taylor & Francis. <http://www.ncbi.nlm.nih.gov/books/NBK56011/>
- Reddy, O. C., & Van Der Werf, Y. D. (2020). The Sleeping Brain: Harnessing the Power of the Glymphatic System through Lifestyle Choices. *Brain Sciences*, 10(11), 868. <https://doi.org/10.3390/brainsci10110868>
- Rosas, N., Sotelo, J., & Nieto, D. (1986). ELISA in the Diagnosis of Neurocysticercosis. *Archives of Neurology*, 43(4), 353–356. <https://doi.org/10.1001/archneur.1986.00520040039016>
- Shenoy, S. S., & Lui, F. (2023). Neuroanatomy, Ventricular System. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK532932/>
- Simon, M. J., & Iliff, J. J. (2016). Regulation of cerebrospinal fluid (CSF) flow in neurodegenerative, neurovascular and neuroinflammatory disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1862(3), 442–451. <https://doi.org/10.1016/j.bbadis.2015.10.014>
- Telano, L. N., & Baker, S. (2023). Physiology, Cerebral Spinal Fluid. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK519007/>
- Veening, J. G., & Barendregt, H. P. (2010). The regulation of brain states by neuroactive substances distributed via the cerebrospinal fluid; a review. *Cerebrospinal Fluid Research*, 7(1), 1. <https://doi.org/10.1186/1743-8454-7-1>
- Ysraelit, M. C., & Correale, J. (2019). Impact of sex hormones on immune function and multiple sclerosis development. *Immunology*, 156(1), 9–22. <https://doi.org/10.1111/imm.13004>
- Zhu, G., Zhou, S., Xu, Y., Gao, R., Zhang, M., Zeng, Q., Su, W., & Wang, R. (2023). Chickenpox and multiple sclerosis: A Mendelian randomization study. *Journal of Medical Virology*, 95(1), e28315. <https://doi.org/10.1002/jmv.28315>

Doping Agents in Sports By ZhongYu Richard Xun

Abstract

Even in 2024, doping in sports remains a significant concern, posing challenges regarding health, ethics, and equity. This paper provides an exploration of doping agents, with a focus on their chemistry, mechanisms of action and methods of their detection. The structures and modes-of-action of testosterone derivatives are discussed, and the analytical methods of gas chromatography coupled with mass spectrometry, gas chromatography coupled with tandem mass spectrometry, gas chromatography/combustion/isotope ratio mass spectrometry, and liquid chromatography are introduced. The paper concludes with a discussion of ways in which users of doping agents attempt to avoid detection and introduces the emerging field of gene doping.

The origin of doping agents

Doping is the act of taking illegal substances to improve physical performance in sports. It has been a persistent issue for decades, undermining the principles of a fair sporting environment; it is used by athletes that wish to improve their performance and gain a competitive advantage. Among the myriad of substances used as doping agents, anabolic androgenic steroids (AASs) are the most popular and common ones used today (Te Kahu Raunui) and will be the focus of this review.

AASs are synthetic derivatives of the natural hormone testosterone (anabolic refers to growing metabolism and androgenic refers to the development of male characteristics). AASs are a type of exogenous steroid—hormones not naturally produced by the body—while testosterone, which naturally occurs in the body, is known as an endogenous steroid. Applications of testosterone date back thousands of years to ancient Greece, where it was first noted that male reproductive organs have healing powers (Chrysopoulos; Dotson and Brown). In addition, the ancient Greeks used plant and testicular extracts that allegedly had performance-enhancing effects (Hoberman and Yesalis). It was not until 1849 that research by Arnold Adolf Berthold proved that castration and reimplantation of testicular tissue had effects on the appearance and behaviour of roosters, thereby proving the existence of a ‘blood stream substance’. He found that roosters had increased combs, interest in hens, and aggressive male behaviour when testicular tissues were reimplanted after castration (Dotson and Brown).

As illustrated by the name, some of the physiological effects of AAS that incentive athletes to abuse them include: increased muscle strength, bone density, and red blood cell production and thus oxygen transport (Baum). Typically, most abusers use steroids from 10 times to 100 times the normal therapeutic doses (Trenton and Currier). These steroids are taken in 4–12-week cycles with a subsequent 4–12-week abstinence period. This is to maximise the drug’s desired effects while minimising its adverse effects (Blue and Lombardo). Many abusers will take steroids through practices known as stacking and pyramiding. Stacking describes the practice where multiple steroids are taken simultaneously, as it is thought that these steroids will have a synergistic effect, despite no scientific evidence to support such assumptions. Pyramiding

is when abusers gradually increase their drug intake until they reach maximum intake during the middle of their cycle, and then gradually taper off their cycle (Trenton and Currier).

Nowadays, steroids are banned from sport competitions in most countries under the laws set by the World Anti-Doping Agency (WADA). These laws aim to promote a fair sporting environment for all athletes, and to protecting athletes from the many health related dangers that come with using steroids (WADA). These dangers range from short-term acne, mood swings, decreased sperm count, and swelling at injection sites to more serious long-term effects on the cardiovascular, reproductive, and metabolic systems, such as fertility and heart problems, kidney failure, tumours in liver, and paranoia (American Addition Centers Editorial Staff). Bond et al. presents some of the adverse effects of steroids on the body (Figure 1).

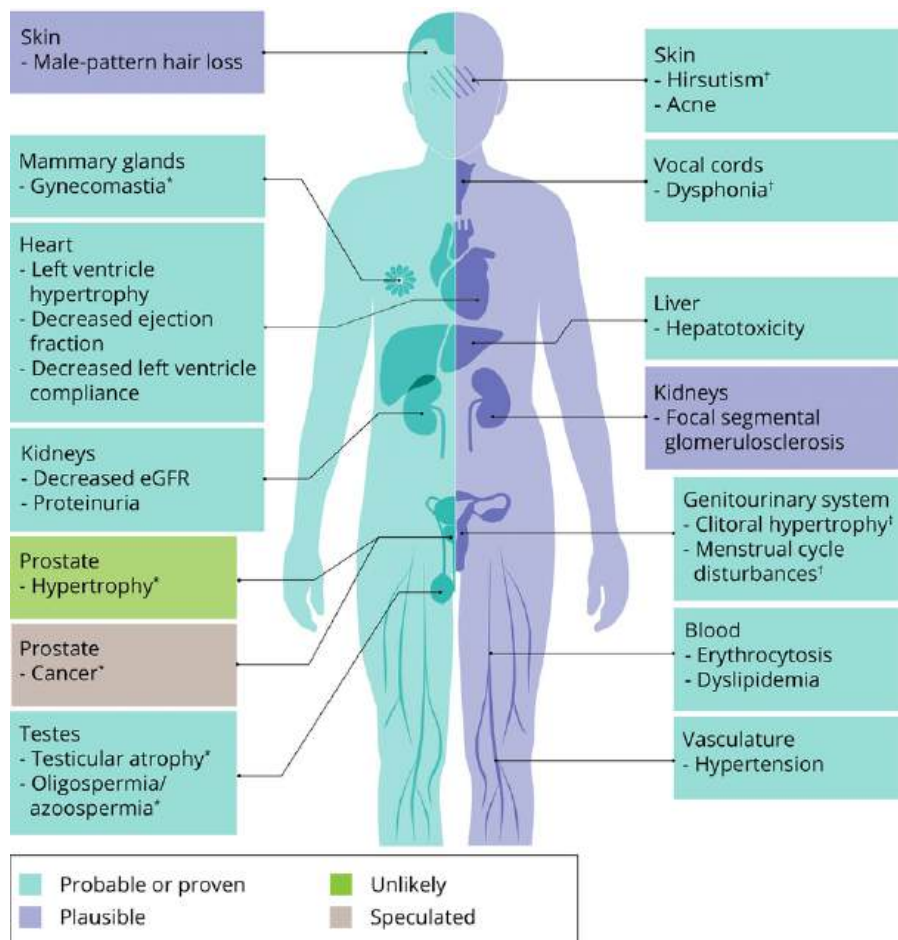


Figure 1: The various adverse effects associated with AAS use (image reproduced from (Bond, Smit, et al.)

A small fraction of officially approved AAS are used medicinally, and are prescribed by doctors, albeit rarely. They are used to treat hormonal diseases, such as delayed puberty; to regain weight after infections, illnesses, or injuries; to treat decreased testosterone in men with medical conditions; to treat certain types of anemia; and to treat hereditary angioedema (Handelsman; Tauchen et al.). AASs come in the form of topical gels, creams, pills, and injectable liquids (Bond, Smit, et al.).

The chemistry and biochemistry of doping agents AAS are mimics of the hormone testosterone. The structures of six of the most commonly abused steroids are shown in Figure 2.

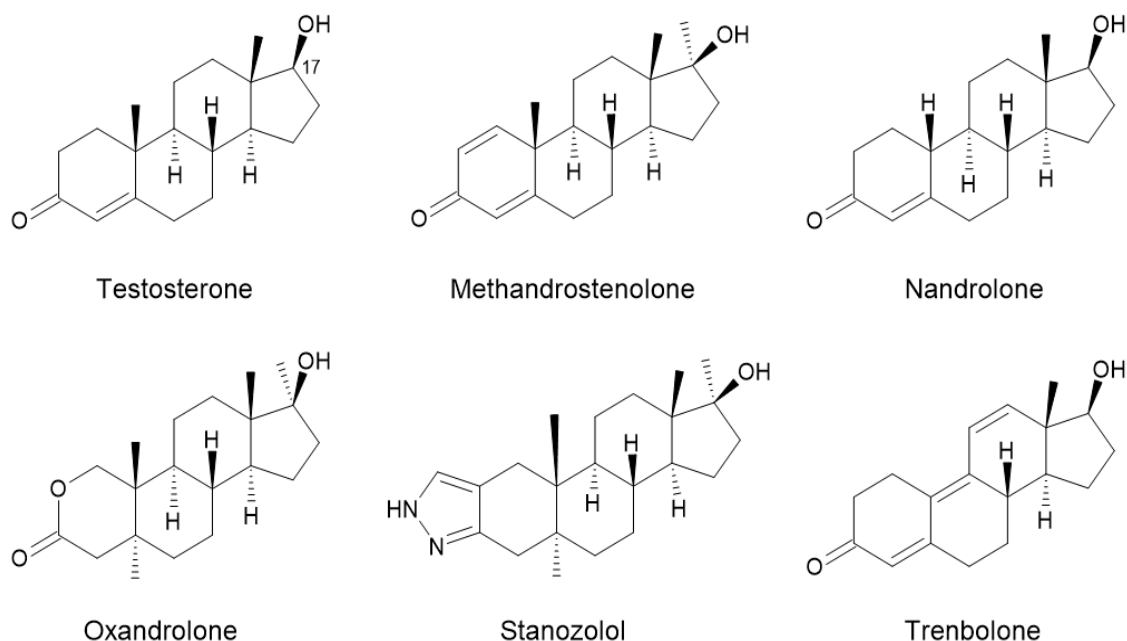


Figure 2: Six of the most commonly abused AAS (drawn with ChemSketch)

As can be seen from Figure 2, all six compounds have a steroid backbone consisting of three six-membered rings and one five-membered ring as well as a 17β-hydroxyl group (illustrated by number 17 on testosterone). The different functional groups of the compounds alter the effects of the drugs. For example, methandrostenolone (Dianabol/D-Bol), oxandrolone, and stanozolol, which are all administered orally, have a 17α-methylation. This modification assist the compounds in resisting hepatic breakdown and increases bioavailability (Bond, Smit, et al.). For steroids administered in an injectable form, a potential modification is the esterification of the 17β-hydroxyl. This modification retards the rate at which the steroid is released from the injected oil-based formulation, thereby prolonging its physiological effect (Bond, Smit, et al.).

The mechanism of action of AASs is analogous to that of testosterone. Testosterone (“testo” meaning testes, “ster” from sterol, and “one” from ketone) is produced naturally in the testes in men and in the ovaries and adrenal glands in women (Davey and Grossmann). Once released into the bloodstream, the majority of circulating testosterone is bound to proteins such as albumin, corticosteroid-binding globulin (CBG) and sex hormone-binding globulin (SHBG), while the rest (typically 1% to 4%) exists in an unbound state (Bond, Llewellyn, et al.; Tan et al.). After diffusion into the cell, testosterone can undergo a variety of biotransformations, either into an androgen with greater potency (such as dihydrotestosterone, which is produced by enzymes in the 5α-reductase family), into a metabolite with less or no potency (by phase I and II metabolism in the liver, kidneys and other androgen-sensitive tissues), or into an estrogen (if testosterone or its derivatives can be substrates for the aromatase enzyme) (Bond, Smit, et al.).

Androgen receptors (ARs) are located in the cytoplasm; when an androgenic hormone, such as testosterone, binds to an AR, a conformational change occurs. Specifically, binding causes the dissociation of heat shock proteins from the AR, which previously had the effect of stabilising the receptor. This is followed by phosphorylation and dimerization of the AR, and, finally, translocation of the AR from the cytoplasm to the nucleus. Once in the nucleus, the AR binds to specific sections of DNA called androgen response elements (AREs) and stimulates the transcription of target genes (Bond, Smit, et al.; Coffey and Robson; Davey and Grossmann).

Methods of detection

There are a variety of factors that determine which method of detection is utilized during the testing process for substance abuse. One such factor is the sensitivity and specificity of the test, which ensures that false positive and false negative results are minimized (Pozo et al.). Having an accurate test decreases the possibility of false accusations, which often come with many legal implications and serious damage to the athlete's reputation. Another factor is a long detection window. All athletes know that they will be tested for steroid use before competitions, thus, it is typical for athletes on steroids to end their cycles long before competition. A long detection window ensures that steroids can be detected well after the athlete stops using them.

For this paper, only methods relating to the detection of steroids will be considered.

Types of biological samples:

Before the testing begins, a type of biological sample must be chosen to be sent to the laboratory; a variety of biological samples are available for the detection of steroids in athletes (Table 1). Some of the most common sample types, ranked in order of increasing detection window include: saliva (up to 24 hours), blood (up to 14 days), urine (up to 28 days), and hair (up to 12 months depending on hair length) (DNA Legal). Saliva tests are not used as standalone methods due to both their short detection window and risk of contamination from food, drink and oral products (*Urine vs Saliva Drug Testing: Pros and Cons - Australia Drug Testing*). However, saliva tests can provide rapid results and are a non-invasive approach. They work because when steroids enter the blood stream, they can diffuse into the salivary glands where they are incorporated with saliva (Lewis). Saliva tests are not often utilized during sporting events. Blood samples are a more invasive approach and require a longer analysis time. Although blood samples also have a relatively short detection period (up to 14 days) compared to other sample types, they are more reliable and provide information on the current circulation of steroids within the bloodstream (DNA Legal). They are typically used for the Athlete's Biological Passport (ABP), which is introduced below, or if urine tests are inconclusive (Alquraini and Auchus). Urine is the traditional type of biological sample used for detection of steroids (Hübener). With a relatively longer detection window (up to 28 days), urine collection is also non-invasive and is more effective and reliable; for most drugs, higher concentrations can be detected in urine compared to that of blood or other sample types (Hübener; Trout and Kazlauskas). Unlike blood, urinal samples provide an averaged presentation of the discarded metabolites from the body.

However, one downside of urine testing is that without supervision, samples can easily be faked and manipulated during collection. Urine tests are typically conducted randomly before the competition or even immediately after the athlete’s event; the test results are added to the ABP (Alquraini and Auchus). Hair testing has the longest detection window of up to 12 months and can provide a month-by-month report of drug usage. Although it is also non-invasive and can be obtained from any part of the body, hair is also commonly subjected to external contamination and it is not suitable in detecting recent use as it takes time for substances to be incorporated in the hair (DNA Legal). Hair testing can still be used if there are suspicions of long-term steroid use (Alquraini and Auchus).

Table 1: Summary of biological sampling types

Type of biological sample	Detection window	Invasiveness	Downsides	How they work
Urine	Up to 28 days	Non-invasive, collected through supervised urine testing	Samples can easily be faked with a lack of supervision	Provide an averaged presentation of the discarded metabolites from the body
Blood	Up to 14 days	Invasive, obtained through blood collection with needles	Can return false negatives due to relatively short detection window.	Presents information regarding the current circulation of steroids in the bloodstream.
Hair	Up to 12 months	Non-invasive, can be collected from any part of the body	Easily subjected to external contamination	Steroids in the bloodstream can become incorporated into hair follicles
Saliva	24 hours	Non-invasive	Risks contamination from food, drinks, oral products	Steroids entering the blood stream can diffuse into the salivary glands

Athlete’s Biological Passport (ABP)

The ABP is an electronic record of professional athletes’ normal biological markers. From these biological markers, an acceptable range is created, and if an athlete’s biological markers exceed this range, it indirectly outlines the possible use of prohibited substances (Krumm et al.). ABP consists of two modules, the first including hematological markers—used to detect blood doping—and the second including urinary markers—used to detect pseudo-endogenous steroids (Krumm et al.). When new samples are recorded, they are passed through The Adaptive Model in ADAMS (Anti-Doping Administration and Management System), which will produce an Atypical Passport Finding (APF) if the likelihood of each marker falling outside the predicted range is lower than 1 in 100 (Goldschmidt). When an APF is flagged, human experts will be informed, and expert evaluation will be conducted.

General method of detecting steroids

The current method of detecting pseudo-endogenous steroids in athletes consists of two different steps: the first step being a longitudinal evaluation by Gas Chromatography coupled

with Mass Spectroscopy (GC/MS), and a confirmation analysis by Gas Chromatography Combustion Isotope Ratio Mass Spectrometry (GC/C/IRMS) (De La Torre et al.).

Gas chromatography coupled with mass spectroscopy (GC/MS)

Gas chromatography was first used in a professional sporting setting at the 1980 Moscow Olympic Games as suggested by Dr Manfred Donike, one of the two heads of the International Olympic Committee Medical and Scientific Department (IOC MSD) (Krieger). At the time, the results showed that more than 20% of samples contained an unnatural ratio of testosterone (Krieger).

GC is a traditional technique that is used to separate organic compounds. It consists of two phases, a mobile phase (an inert carrier gas such as N₂, He or Ar) and a stationary phase (a packed column that is coated with a high boiling polymer such as silica gel) (UCLA). The analyte is introduced before the column, and is separated based on the interactions that it has with the stationary phase; the stronger the interaction is, the longer it takes to migrate through the column and therefore the longer its retention time (UCLA). This separation can occur by gas-liquid partition or gas-solid adsorption, based on the type of stationary phase used (Arruda et al.). A gas chromatogram is drawn, where the peak position represents the retention time of the sample and the peak area is proportional to the quantity of sample.

After exiting the column, the samples are analysed using quadrupole mass spectroscopy, where ionised samples are focused and passed along the middle of the quadrupole rods. These rods have fixed direct current (DC) and alternating radiofrequency potentials (RF) potentials applied to them; by varying the DC and RF values, different ions of different mass-to-charge ratios (m/z) are filtered (Tyagi et al.). A mass spectrum can be obtained in three forms based on which mode the detector was set too. Full scans—where the entire mass range for the fragment ions is scanned—and selected ion monitoring (SIM)—where specific ions are chosen to be scanned (Snow).

The information gathered from GC/MS can then be compared with existing data from known metabolites, thus indicating the presence of steroids within the athlete. Alternatively, the quantity of testosterone in the body can be compared to its isomer epitestosterone in the testosterone/epitestosterone ratio (T/E). This is because the isomer epitestosterone is typically found in the body in similar quantities of testosterone; when athletes dope, their testosterone levels will increase greatly while their epitestosterone quantity remains relatively stable (Danaceau et al.). WADA has set a 4.0 T/E ratio as an indicator for possible exogenous use. When athlete's urine surpasses this level, subsequent analysis is conducted with GC/C/IRMS (Danaceau et al.).

Sample preparation

Before GC/MS can be conducted, the urine sample must be prepared for GC/MS. Sample preparation is conducted with the aim of making the sample more suitable for the chromatographic environment while also maintaining the sample's integrity as much as possible

(Fernández-Peralbo and Luque De Castro). Sample preparation of urine for GC/MS analysis typically entails the following steps (Figure 3).

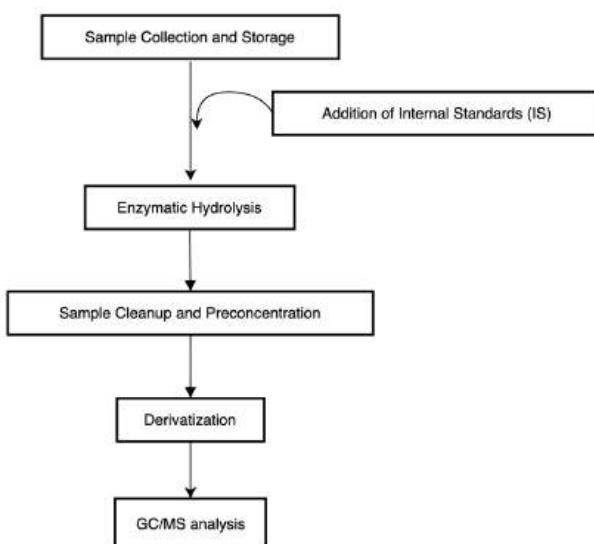


Figure 3: An illustration of sample preparation for GC/MS analysis

During sample collection and storage, urine samples should be frozen at temperatures around $-80\text{ }^{\circ}\text{C}$ to ensure the targeted metabolites within the sample are preserved and stable; however, the temperature can be adjusted specifically for a known metabolite. (Fernández-Peralbo and Luque De Castro).

Internal Standards (IS)

In the laboratory, the sample undergoes treatment in preparation for GC/MS analysis. One of the primary reasons for this is to decrease signal suppression from other matrix components and to make it easier to read the data produced by GC/MS. To start off, an IS is chosen and a known quantity is added to the sample. The sample is then calibrated, and the peak area of the drug is compared to that of the internal standard. Internal standards are compounds that have almost identical chemical properties to the analyte, typically isotopically labelled analogues of the target compounds (Mischak et al.). Common ISs are deuterated forms of the target compounds or ones labelled with ^{13}C ; these are known as stable isotope labelled internal standards (SIL IS) (Fernández-Peralbo and Luque De Castro). They are used to calibrate the data.

One of the main purposes of ISs is quantification. The area under the peak of an analyte or an IS is proportional to the concentration of that compound. By adding known quantities of IS, the concentration of the analyte can be calculated. The different detection efficiencies of the analyte versus the IS can be corrected for via the calculation of a response factor (the ratio of peaks areas of analyte/IS when equal concentrations of the two compounds are added). This

allows for quantification of the drug, despite potential errors in volumetric recovery or amount of injection solvent added (SeparationScience).

Another reason for using ISs is to account for signal suppression or enhancements between the analyte and matrix components (Fernández-Peralbo and Luque De Castro). Suppression can occur due to coelution between endogenous substances in the urine and analytes. Coelution is when multiple compounds elute from the chromatography column at the same time, and compete with the analyte for the total available charge in the MS detector. This can weaken the signal and reduce the accuracy of the results by as much as 26% (Hewavitharana et al.). ISs are often coeluted with the analyte and are added first before treatment so that they are subjected to the same treatment process as the analytes. This means that these ISs will be able to provide a consistent reference to the entire analytical process. As a result, the analyte peak response is usually corrected for the ion suppression matrix effects (Hewavitharana et al.).

Enzymatic hydrolysis

After the addition of an IS, enzymatic hydrolysis is conducted. Before excretion of the target compounds from the body, the compounds are typically converted into more hydrophilic products such as conjugated metabolites (Fernández-Peralbo and Luque De Castro). These conjugated forms (commonly sulfate and glucuronide derivatives) are more difficult to detect and quantify with GC/MS and other analytical tests (Allende et al.). To counteract this, enzymatic hydrolysis reverts conjugated metabolites back into their free steroid form. Some of the common biocatalysts used in the preparation of urinary samples for steroid detection include β -glucuronidase (for hydrolysing glucuronide conjugates), arylsulfatase (for hydrolysing sulfate conjugates), or microbial enzymes, such as those derived from *Cunninghamella blakesleeana* (Skov et al.).

Sample cleanup and preconcentration

Sample cleanup and preconcentration removes interfering substances while concentrating the analytes for better sensitivity and results. The most common procedure is solid-phase extraction (SPE). One typical method for SPE uses syringe-shaped SPE tubes, containing the solid phase—such as C₁₈, C₅, or C₈ for water rich urine samples—kept in place by two filters (Stevenson). When a urine sample is applied and washed through the solid phase with a wash solvent, the metabolites bind to the solid phase due to hydrophilic, hydrophobic, or ionic interactions (Solid Phase Extraction SPE; Mandal et al.). Matrix components are washed through as waste. Then, an elution solvent strong enough to extract the analyte is applied and the analyte is collected (Solid Phase Extraction SPE; Mandal et al.).

Derivatization

Derivatization is the process of converting the chemical structure of a compound to a structure that has better analytical capabilities. Types of derivatization include silylation (using trimethylchlorosilane (TMCS) to replace the active hydrogens in the functional groups of the

compound with trimethylsilyl (TMS)), acylation (using acetic anhydride or trifluoroacetic anhydride (TFAA) to introduce an acyl group to compounds containing active hydrogen groups), and alkylation (using methyl iodide or dimethyl sulfate to replace the active hydrogen groups with alkyl groups) (Lord and Pfannkoch; Lin et al.; Sigma-Aldrich Co).

Ideally, if the analyte can be tested in its original form, it should be. Derivatization of an analyte into another form adds costs, uncertainties and impurities. However, there are reasons why derivatization might be used for an analyte. Lin et al. suggests three reasons including: to make the analytes more chemically compatible with the chromatographic environment, to achieve required separation or improve separation efficiency, and to improve detection effectiveness. An example where derivatization is used to improve analyte compatibility with the chromatographic environment is when analytes have carboxyl or amine functional groups, which can form hydrogen-bonds with the chromatography system. The formation of hydrogen bonds with the chromatography system results in peak loss by irreversible adsorption and peak tailing by reversible adsorption. Derivatization can change these species into their inactive form prior to GC analysis and prevent hydrogen bonding.

An example where derivatization is used to achieve the required separation for MS analysis is for enantiomers; typical GC cannot separate enantiomers due to their identical physical properties. Derivatization with chiral reagents can be used to separate enantiomers by conversion to diastereomers. Diastereomers have different physical properties and can be separated by GC. Derivatization is often preferred over the use of chiral GC columns—which also separate enantiomers—due to its cost-effective nature (Bogos et al.).

Gas chromatography paired with tandem mass spectroscopy GC/MS/MS

There are many types of GC/MS/MS, including GC-TQMS (triple quadrupole), GC-ITMS (ion trap MS) and GC-QTOF-MS (quadrupole time of flight). The most commonly used in the detection of steroids from urinary samples is GC-TQMS (Snow). The main difference between GC/MS and GC-TQMS is the presence of three quadrupole mass filters on GC-TQMS compared to one on GC/MS. This allows for fragmentation, which enhances the effectiveness of the method by reducing noise and increasing specificity and sensitivity (Turner).

The first of these mass filters operates identically to that in GC/MS, it filters ions that will be scanned. The second quadrupole acts as a medium for collision induced fragmentation, creating new fragments that are passed to the third quadrupole, while filters and analyses them (Snow). GC-TQMS is powerful in this way, larger fragments can be broken down to confirm the correct structure of the molecule. Another reason why GC-TQMS is known as a “multidimensional” technique is that all of the quadrupoles can be set for different functions. For instance, all three quadrupoles can be set to the function of the first quadrupole, which achieves a traditional GC/MS analysis. Alternatively, the first quadrupole can be set to SIM for a single ion and third quadrupoles can be set to SIM for a single fragment from that ion. This allows for ultimate sensitivity; it is very unlikely for an ion (or even isomer) to have the same transition after being fragmented and thus the same retention time. One of the disadvantages of using GC-TQMS is the

cost of the machine and the special training required to use the machine. Additionally, errors in sample preparation and contamination can be amplified when using GC-TQMS.

Gas chromatography/combustion/paired with isotopic ratio mass spectroscopy (GC/C/IRMS)

GC/MS is used in the initial stages of detection due to its fast and cost-effective nature. On the other hand, GC/MS cannot differentiate between endogenous steroids and synthetic analogs (Van Renterghem et al.). However, GC/C/IRMS can, and is used as a confirmation procedure after anomalies are detected using GC/MS. GC/C/IRMS works on the principle that the relative ratio of light stable isotopes in natural compounds are different to those in synthetic compounds (Hackney). For example, De La Torre et al. found that the $^{13}\text{C}/^{12}\text{C}$ isotopic ratio for natural human testosterone deviated from the $^{13}\text{C}/^{12}\text{C}$ ratio of a defined standard by -21.3‰ to -24.4‰ while synthetic testosterone had deviations ranging from -26.18‰ to -30.04‰. GC/C/IRMS rules out false positives that could potentially be simply variations in the athlete's endogenous steroid profile. Some of the isotopes that GC/C/IRMS looks at includes carbon isotopes ($^{13}\text{C}/^{12}\text{C}$), hydrogen isotopes ($^2\text{H}/^1\text{H}$), nitrogen isotopes ($^{15}\text{N}/^{14}\text{N}$), and oxygen isotopes ($^{18}\text{O}/^{16}\text{O}$). However, GC/C/IRMS faces difficulties when the steroid is purposefully created with a similar carbon isotopic composition value to those reported for endogenous urinary steroids.

Liquid chromatography tandem mass spectrometry

GC/MS has a multitude of limitations including being unable to detect thermally unstable compounds, non-volatile compounds, and polar compounds. This is when an alternative technique—liquid chromatography coupled with mass spectrometry or tandem mass spectrometry—can be used. Liquid chromatography (LC) is a similar chromatographic technique to GC; however, a liquid mobile phase is used instead of a gaseous mobile phase. Unlike GC, where separation is mainly determined by the boiling points of the solute molecules, LC separation is determined by the interaction of the solute with the chromatography medium (Cheriyedath). Although slower, LC is used when the solute is thermally unstable and could be structurally altered when placed in the high temperature GC column. The technique also requires less sample preparation and virtually no derivatisation, while being a more sensitive chromatographic technique in comparison to GC/MC (A. Schug).

Preventing detection

Masking agents are a group of compounds that athletes use to prevent the detection of steroids. There are many types of masking agents including diuretics, which change the composition of body fluids by increasing the rate of urine flow (Cadwallader et al.); probenecid, which reduces the secretion of steroids in the urine (Hemmersbach); and liposomes, which work by encapsulating the drugs and modifying its physiochemical and pharmacokinetic properties (Botrè). Both diuretics and probenecid can be detected by GC/MS or LC/MS (Cadwallader et al.); however, liposomes cannot. Liposomes are aqueous core filled spheres surrounded by a

phospholipid bilayer; they can be mixed with steroids during injection, injected just before an expected doping test, or added into the urine sample during collection (Botrè). There are two ways which liposomes mask the presence of steroids: “body orientated” masking, meaning that they slow down the release of the drug altering its pharmacokinetics, and “lab-orientated” meaning that they interfere with analytical methods by creating a lipid drug complex with free steroids/metabolites, decreasing the effectiveness of standardised tests such as GC and LC (Schänzer et al.).

Designer Drugs

Another way in which athletes avoid detection of doping is through designer drugs. Designer drugs are drugs that have been structurally manipulated to avoid detection in WADA accredited laboratories. Because of a lack of research surrounding these drugs, designer drugs are consequently far more dangerous than typical drugs (Kazlauskas). The development of detection for designer drugs is an evolving area of research. This is discussed in more detail below.

Microdosing

Microdosing is the technique of taking small doses of doping agents in just the right amount. It takes advantage of WADA’s T/E ratio, which allows for a testosterone to epitestosterone ratio of 4.0. By taking micro doses, athletes can stay within the allowed range of T/E ratio fluctuation, while still achieving noticeable benefits that improve performance. The idea behind microdosing is that the athlete has the benefits of the doping agents during the event, yet evidence of its presence dissipates before the athlete is tested afterwards (Brodwin).

New detection techniques

Currently, new techniques are being created in an attempt to improve detection of doping agents and decrease the possibilities of false positives. Two methods that are able to detect the use of liposomes as masking agents are currently under investigation: one being flow cytometry, used to detect liposomes in the blood, and the other being detection—by LC/MS—of DSPE-PEG, a product of liposome breakdown present in urine (Schänzer et al.).

Another emerging technique that allows for the detection of steroids is the bioassay. This is the process of measuring the potency of a drug by the effect it has on living organisms or tissues. Bioassays act as a non-targeted approach in detecting doping substances, including designer drugs (Chayen and Bitensky). Bioassays work on the principle that all steroids have the same mechanism of action in the process of asserting their effects. One common form of bioassay uses cultured cells that contain an AR in their cytoplasm. When an androgen binds to the receptor, it releases the protein complex HSP90 which then dimerises a second ligand-bound AR. The second ligand-bound AR translocates to the nucleus, augmenting DNA transcription by binding with androgen receptor elements (ARE). Bioassays take cells from different sources (e.g. HuH7 (human liver cancer cells) or MDA-kb2 (human breast cancer cells)) and genetically modify them to be able to express reporter proteins under the regulation of AREs (Lund et al.).

One specific bioassay proven to be very effective in the detection of doping substances in urine is the Chemically Activated Luciferase eXpression (CALUX) bioassay, or AR-CALUX, when applied to androgens (Martín-Escudero et al.). AR-CALUX works by incorporating androgen-controlled luciferase reporter gene constructs into human U2-OS cells. When androgens enter the cells, transcription occurs and the firefly gene is produced, ultimately emitting light; luciferase activity can then be measured using a luminometer. This process is shown in more detail in Figure 4.

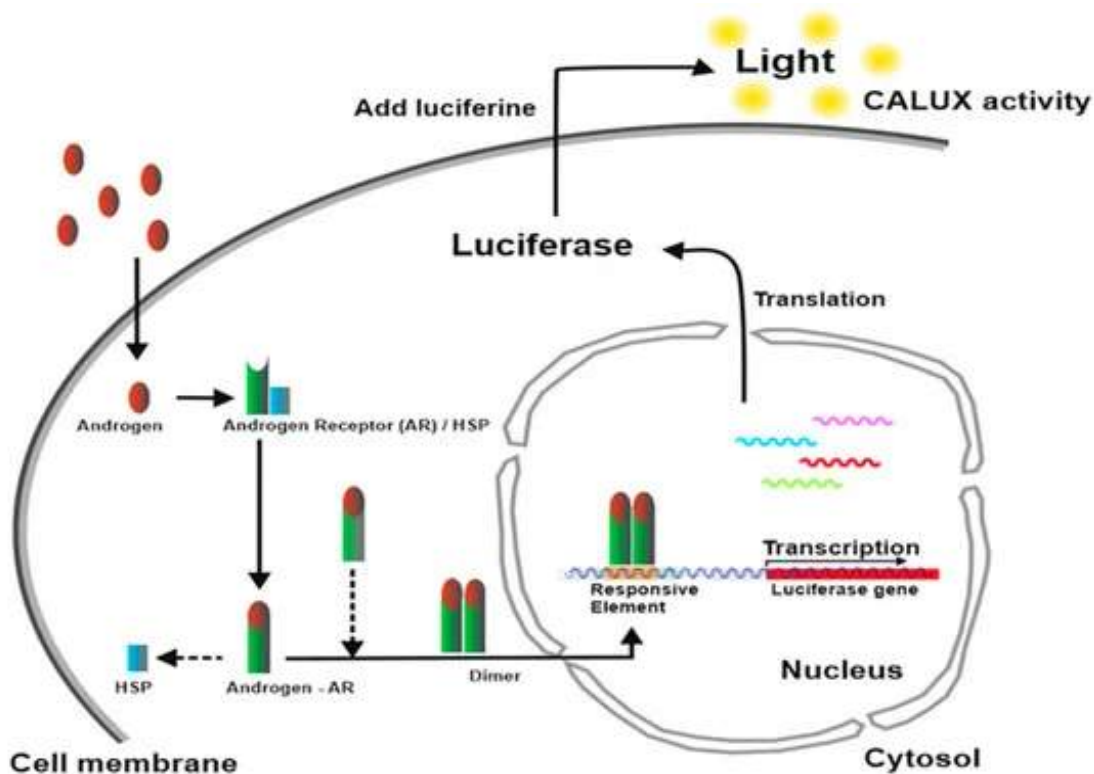


Figure 4: the process of AR-CALUX bioassays (image reproduced from (Martín-Escudero et al., 2021))

One of the reasons why some cell lines are preferred over others is a lack of other receptors. The absence of certain receptors means that there is less chance of crosstalk between receptors, a phenomenon where similar steroid/androgen receptors recognise the same hormone response element, leading to the presence of false positives (Lund et al.). Different cell types also have different biological mechanisms. Yeast cells do not detect designer androgens in the form of prohormones due to their lack of steroid-metabolizing enzymes, but are better at measuring intrinsic androgenic potential due to this fact (Lund et al.). Mammalian cells also have steroid metabolizing enzymes; however, each type of mammalian cell has different expression patterns.

Another recent technique for steroid detection is nano liquid chromatography mass spectroscopy (nLC/MS). Reducing the column internal diameter (to < 0.5 mm) in nLC can reduce sample and solvent demands while also enhancing sensitivity with MS (Sanders and Edwards). Lower detection limits achieved through nLC/MS will help identify use of doping substances easier. It was also found that this process could be automated (Thevis).

The future of doping

With the evolution of detection methods for steroids, scientists have also begun to explore and prepare for alternative methods that could be used by athletes in the future; one of which is gene doping. Gene doping is the process of employing gene editing techniques—including modifying the human genome or to introduce transgenes—to improve the sporting performances of athletes (Li et al.). Although no athletes have been caught doing gene doping, scientists are fully aware of its possibility and thus are still improving detection methods for it (Proszenko).

Currently, some genes that could potentially improve athlete performances have been identified. This includes the *PPARGC1A* gene, which can significantly improve endurance by enhancing mitochondrial function (Lippi et al.; Varillas-Delgado et al.); the T allele on the *AMPD1* gene, which improves skeletal muscle energy metabolism (Varillas-Delgado et al.); and the *COL5A1* gene, which is correlated with ligament injuries and thus, potentially recovery times for athletes (Alvarez-Romero et al.; Posthumus et al.). Because gene doping is a relatively novel technique, its long-term effects on athlete health have not yet been established (Li et al.). However, there are also an increasing number of methods that can be used in the detection of gene doping of which include polymerase chain reaction (PCR) techniques, MS techniques, and CRISPR-based techniques.

Conclusion

The everlasting race between athletes that abuse doping agents and antidoping agencies is often compared to that of a cat and mouse chase. The development of new drugs is always met with the development of new detection methods. As long as there are incentives for performance enhancement, athletes will continue to seek ways to abuse certain substances. GC/MS, GC/MS/MS, LC/MS, LC/MS/MS, nLC/MS and bioassays are all integral techniques in the detection of steroids; their refinement disincentives athletes from pursuing unethical doping practices. Furthermore, new methods to counter gene doping are in development, with hopes of detecting gene doping even before its abuse. Understanding the chemistry and biology of steroid detection is necessary in the process of upholding a fair sporting ground for all.

Works Cited

- A. Schug, Kevin. The LCGC Blog: Forensic Drug Analysis: GC–MS versus LC–MS. 5 July 2018, <https://www.chromatographyonline.com/view/lcgc-blog-forensic-drug-analysis-gc-ms-versus-lc-ms>.
- Allende, Fidel, et al. “LC–MS/MS Method for the Simultaneous Determination of Free Urinary Steroids.” *Chromatographia*, vol. 77, no. 7–8, Apr. 2014, pp. 637–42. DOI.org (Crossref), <https://doi.org/10.1007/s10337-014-2638-4>.
- Alquraini, Hussain, and Richard J. Auchus. “Strategies That Athletes Use to Avoid Detection of Androgenic-Anabolic Steroid Doping and Sanctions.” *Molecular and Cellular Endocrinology*, vol. 464, Mar. 2018, pp. 28–33. DOI.org (Crossref), <https://doi.org/10.1016/j.mce.2017.01.028>.
- Alvarez-Romero, Javier, et al. “Genetic Variants within the COL5A1 Gene Are Associated with Ligament Injuries in Physically Active Populations from Australia, South Africa, and Japan.” *European Journal of Sport Science*, vol. 23, no. 2, Feb. 2023, pp. 284–93. DOI.org (Crossref), <https://doi.org/10.1080/17461391.2021.2011426>.
- American Addiction Centers Editorial Staff. “Effects of Steroids | Short & Long-Term Effects of Steroid Use.” *DrugAbuse.Com*, 11 Mar. 2024, <https://drugabuse.com/stimulants/steroids/effects-use/>.
- Arruda, Tarsila Rodrigues, et al. “Emerging Techniques for Extraction and Characterization of Natural Compounds.” *Green Products in Food Safety*, Elsevier, 2023, pp. 29–79. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-323-95590-4.00009-6>.
- Baum, Zach. “The Science behind Performance-Enhancing Drugs.” CAS, 21 July 2021, <https://www.cas.org/fr/resources/cas-insights/science-behind-performance-enhancing-drugs>.
- Blue, Jeffrey G., and John A. Lombardo. “STEROIDS AND STEROID-LIKE COMPOUNDS.” *Clinics in Sports Medicine*, vol. 18, no. 3, July 1999, pp. 667–89. DOI.org (Crossref), [https://doi.org/10.1016/S0278-5919\(05\)70175-7](https://doi.org/10.1016/S0278-5919(05)70175-7).
- Bogos, Luisa-Gabriela, et al. “Indirect Enantioseparations: Recent Advances in Chiral Metabolomics for Biomedical Research.” *International Journal of Molecular Sciences*, vol. 23, no. 13, July 2022, p. 7428. PubMed, <https://doi.org/10.3390/ijms23137428>.
- Bond, Peter, William Llewellyn, et al. “Anabolic Androgenic Steroid-Induced Hepatotoxicity.” *Medical Hypotheses*, vol. 93, Aug. 2016, pp. 150–53. DOI.org (Crossref), <https://doi.org/10.1016/j.mehy.2016.06.004>.
- Bond, Peter, Diederik L. Smit, et al. “Anabolic–Androgenic Steroids: How Do They Work and What Are the Risks?” *Frontiers in Endocrinology*, vol. 13, Dec. 2022. Frontiers, <https://doi.org/10.3389/fendo.2022.1059473>.
- Botrè, F. Are Liposomes Masking Agents? An in-Progress Study. 2009.
- Brodwin, Erin. “Olympians May Be Taking Cues from Silicon Valley’s Favorite Way to Do Drugs.” *BusinessInsider*, 13 Feb. 2018, <https://www.businessinsider.com/doping-cheating-winter-olympics-microdosing-testosterone-steroids-2018-2>.
- Cadwallader, Amy B., et al. “The Abuse of Diuretics as Performance-Enhancing Drugs and Masking Agents in Sport Doping: Pharmacology, Toxicology and Analysis.” *British Journal of Pharmacology*, vol. 161, no. 1, Sept. 2010, pp. 1–16. PubMed, <https://doi.org/10.1111/j.1476-5381.2010.00789.x>.

- Chayen, J., and L. Bitensky. "BIOASSAYS | Overview." *Encyclopedia of Analytical Science*, Elsevier, 2005, pp. 259–65. DOI.org (Crossref), <https://doi.org/10.1016/B0-12-369397-7/00041-8>.
- Cheriyedath, Susha. "Liquid Chromatography versus Gas Chromatography." *New Medical Life Sciences*, <https://www.news-medical.net/life-sciences/Liquid-Chromatography-versus-Gas-Chromatography.aspx>. Accessed 8 Mar. 2024.
- Chrysopoulos, Philip. "Doping in Ancient Greece." 21 Oct. 2022, <https://greekreporter.com/2022/10/21/doping-sports-ancient-greece/>.
- Coffey, Kelly, and Craig N. Robson. "Regulation of the Androgen Receptor by Post-Translational Modifications." *Journal of Endocrinology*, vol. 215, no. 2, Nov. 2012, pp. 221–37. DOI.org (Crossref), <https://doi.org/10.1530/JOE-12-0238>.
- Danaceau, Jonathan P., et al. "Quantitative Confirmation of Testosterone and Epitestosterone in Human Urine by LC/Q-ToF Mass Spectrometry for Doping Control." *Journal of Mass Spectrometry*, vol. 43, no. 7, July 2008, pp. 993–1000. DOI.org (Crossref), <https://doi.org/10.1002/jms.1443>.
- Davey, Rachel A., and Mathis Grossmann. "Androgen Receptor Structure, Function and Biology: From Bench to Bedside." *The Clinical Biochemist. Reviews*, vol. 37, no. 1, Feb. 2016, pp. 3–15.
- De La Torre, X, et al. "¹³C/¹²C Isotope Ratio MS Analysis of Testosterone, in Chemicals and Pharmaceutical Preparations." *Journal of Pharmaceutical and Biomedical Analysis*, vol. 24, no. 4, Feb. 2001, pp. 645–50. DOI.org (Crossref), [https://doi.org/10.1016/S0731-7085\(00\)00452-0](https://doi.org/10.1016/S0731-7085(00)00452-0).
- De La Torre, Xavier, et al. "Improving the Detection of Anabolic Steroid Esters in Human Serum by LC-MS." *Journal of Pharmaceutical and Biomedical Analysis*, vol. 194, Feb. 2021, p. 113807. DOI.org (Crossref), <https://doi.org/10.1016/j.jpba.2020.113807>.
- DNALegal. "A Guide to Anabolic Steroids Drug Testing." DNALegal Helping You Find the Truth, 22 Apr. 2024, <https://www.dnalegal.com/anabolic-steroids-drug-testing>.
- Dotson, Jennifer L., and Robert T. Brown. "The History of the Development of Anabolic-Androgenic Steroids." *Pediatric Clinics of North America*, vol. 54, no. 4, Aug. 2007, pp. 761–69. DOI.org (Crossref), <https://doi.org/10.1016/j.pcl.2007.04.003>.
- Fernández-Peralbo, M. A., and M. D. Luque De Castro. "Preparation of Urine Samples Prior to Targeted or Untargeted Metabolomics Mass-Spectrometry Analysis." *TrAC Trends in Analytical Chemistry*, vol. 41, Dec. 2012, pp. 75–85. DOI.org (Crossref), <https://doi.org/10.1016/j.trac.2012.08.011>.
- Goldschmidt, Henry. "Demystifying the Athlete Biological Passport." *Squire Patton Boggs*, 14 Feb. 2024, https://www.sports.legal/2024/02/demystifying-the-athlete-biological-passport/?__cf_chl_tk=YXGj.ZIMdyt oJR3hoEJl83BNPDd0M1PkXmpf6MjvBM0-1721347094-0.0.1.1-3966.
- Hackney, Anthony C. "Athlete Testing, Analytical Procedures, and Adverse Analytical Findings." *Doping, Performance Enhancing Drugs, and Hormones in Sport*, Elsevier, 2018, pp. 113–27. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-12-813442-9.00010-9>.

- Handelsman, David J. “Androgen Physiology, Pharmacology, Use and Misuse.” Endotext, edited by Kenneth R. Feingold et al., MDText.com, Inc., 2000. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK279000/>.
- Hemmersbach, Peter. “The Probenecid-Story - A Success in the Fight against Doping through out-of-Competition Testing.” *Drug Testing and Analysis*, vol. 12, no. 5, May 2020, pp. 589–94. PubMed, <https://doi.org/10.1002/dta.2727>.
- Hewavitharana, Amitha K., et al. “The Importance of Complete Overlapping of Analyte and Internal Standard Peaks in Eliminating Matrix Effects with Liquid Chromatography–Mass Spectrometry (LC–MS).” *Intellisphere, LLC*, vol. 39, no. 7, July 2021, https://go-gale-com.ezproxy.kotui.ac.nz/ps/retrieve.do?tabID=T002&resultListType=RESULT_LIST&searchResultsType=SingleTab&retrievalId=43ae3c65-8272-46f4-9612-0e27b96a39fb&hitCount=2&searchType=BasicSearchForm¤tPosition=2&docId=GALE%7CA677209702&docType=Article&sort=Relevance&contentSegment=ZGPP-MOD1&prodId=GPS&pageNum=1&contentSet=GALE%7CA677209702&searchId=R1&userGroupName=per_k12&inPS=true.
- Hoberman, John M., and Charles E. Yesalis. “The History of Synthetic Testosterone.” *Scientific American*, vol. 272, no. 2, Feb. 1995, pp. 76–81. DOI.org (Crossref), <https://doi.org/10.1038/scientificamerican0295-76>.
- Hübener, Hans Joachim. “Steroid Alcohols in Urine.” *Methods of Enzymatic Analysis*, Elsevier, 1965, pp. 485–90. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-12-395630-9.50100-6>.
- Kazlauskas, Ray. “Designer Steroids.” *Doping in Sports*, edited by Detlef Thieme and Peter Hemmersbach, vol. 195, Springer Berlin Heidelberg, 2009, pp. 155–85. DOI.org (Crossref), https://doi.org/10.1007/978-3-540-79088-4_7.
- Krieger, Jörg. “Intended. Underrated. Disputed. The IOC Medical Commission’s ‘Subcommission on Doping and Biochemistry in Sport’ between 1980 and 1988.” *Performance Enhancement & Health*, vol. 4, no. 3–4, June 2016, pp. 88–93. DOI.org (Crossref), <https://doi.org/10.1016/j.peh.2016.03.002>.
- Krumm, Bastien, et al. “Future Opportunities for the Athlete Biological Passport.” *Frontiers in Sports and Active Living*, vol. 4, 2022, p. 986875. PubMed, <https://doi.org/10.3389/fspor.2022.986875>.
- Lewis, John G. “Steroid Analysis in Saliva: An Overview.” *The Clinical Biochemist. Reviews*, vol. 27, no. 3, Aug. 2006, pp. 139–46.
- Li, Ruihong, et al. “Gene Doping Detection in the Era of Genomics.” *Drug Testing and Analysis*, Feb. 2024, p. dta.3664. DOI.org (Crossref), <https://doi.org/10.1002/dta.3664>.
- Lin, D. L., et al. “Chemical Derivatization for the Analysis of Drugs by GC-MS - A Conceptual Review.” *Journal of Food and Drug Analysis*, vol. 16, no. 1, July 2020. DOI.org (Crossref), <https://doi.org/10.38212/2224-6614.2373>.
- Lippi, G., et al. “Genetics and Sports.” *British Medical Bulletin*, vol. 93, no. 1, Mar. 2010, pp. 27–47. DOI.org (Crossref), <https://doi.org/10.1093/bmb/ldp007>.
- Lord, H. L., and E. A. Pfannkoch. “Sample Preparation Automation for GC Injection.” *Comprehensive Sampling and Sample Preparation*, Elsevier, 2012, pp. 597–612. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-12-381373-2.00061-2>.

- Lund, Rachel A., et al. "Nontargeted Detection of Designer Androgens: Underestimated Role of in Vitro Bioassays." *Drug Testing and Analysis*, vol. 13, no. 5, May 2021, pp. 894–902. DOI.org (Crossref), <https://doi.org/10.1002/dta.3049>.
- Mandal, Swagata, et al. "Review of Extraction and Detection Techniques for the Analysis of Pesticide Residues in Fruits to Evaluate Food Safety and Make Legislative Decisions: Challenges and Anticipations." *Journal of Chromatography B*, vol. 1215, Jan. 2023, p. 123587. DOI.org (Crossref), <https://doi.org/10.1016/j.jchromb.2022.123587>.
- Martín-Escudero, Pilar, et al. "Bioanalytical Detection of Steroid Abuse in Sports Based on the Androgenic Activity Measurement." *Chemosensors*, vol. 9, no. 4, Mar. 2021, p. 62. DOI.org (Crossref), <https://doi.org/10.3390/chemosensors9040062>.
- Mischak, Harald, et al. "Comprehensive Human Urine Standards for Comparability and Standardization in Clinical Proteome Analysis." *PROTEOMICS – Clinical Applications*, vol. 4, no. 4, Apr. 2010, pp. 464–78. DOI.org (Crossref), <https://doi.org/10.1002/prca.200900189>.
- Posthumus, Michael, et al. "The COL5A1 Gene: A Novel Marker of Endurance Running Performance." *Medicine & Science in Sports & Exercise*, vol. 43, no. 4, Apr. 2011, pp. 584–89. DOI.org (Crossref), <https://doi.org/10.1249/MSS.0b013e3181f34f4d>.
- Pozo, Oscar J., et al. "Detection and Characterization of Anabolic Steroids in Doping Analysis by LC-MS." *TrAC Trends in Analytical Chemistry*, vol. 27, no. 8, Sept. 2008, pp. 657–71. DOI.org (Crossref), <https://doi.org/10.1016/j.trac.2008.06.003>.
- Prosenko, Adrian. Why Gene Doping Is the next Great Olympic Threat. 20 July 2024, <https://www.smh.com.au/sport/why-gene-doping-is-the-next-great-olympic-threat-20240720-p5jv5w.html>.
- Sanders, Katherine L., and James L. Edwards. "Nano-Liquid Chromatography-Mass Spectrometry and Recent Applications in Omics Investigations." *Analytical Methods*, vol. 12, no. 36, 2020, pp. 4404–17. DOI.org (Crossref), <https://doi.org/10.1039/D0AY01194K>.
- Schänzer, Wilhelm, et al. *Proceedings of the Manfred-Donike-Workshop, 29th Cologne Workshop on Dope Analysis 13th to 18th February 2011*. Sportverl. Strauß, 2012.
- SeparationScience. Internal Standards: How Does It Work? https://learning.sepscience.com/hubfs/Technical%20Blogs/HPLC_134_v2.pdf. Accessed 22 Aug. 2024.
- Sigma-Aldrich Co. Guide to Derivatization Reagents for GC. Supelco, 1997, <https://gcms.cz/labrulez-bucket-strap-i-h3hsga3/application::paper.paper/t196909.pdf>.
- Skov, Kathrine, et al. "Exploring Enzymatic Hydrolysis of Urine Samples for Investigation of Drugs Associated with Drug-Facilitated Sexual Assault." *Pharmaceuticals*, vol. 17, no. 1, Dec. 2023, p. 13. DOI.org (Crossref), <https://doi.org/10.3390/ph17010013>.
- Snow, Nicholas. "Flying High with Sensitivity and Selectivity: GC–MS to GC–MS/MS." *LCGC North America*, vol. 39, no. 2, Feb. 2021, <https://doi.org/10.56530/lcgc.na.yn3065q6>.
- Solid Phase Extraction SPE. Directed by Anneli science blog, 2021. YouTube, https://www.youtube.com/watch?v=yCD_z-q_6iA.
- Stevenson, D. IMMUNOAFFINITY EXTRACTION.
- Tan, Mh Eileen, et al. "Androgen Receptor: Structure, Role in Prostate Cancer and Drug Discovery." *Acta Pharmacologica Sinica*, vol. 36, no. 1, Jan. 2015, pp. 3–23. DOI.org (Crossref), <https://doi.org/10.1038/aps.2014.18>.

- Tauchen, Jan, et al. "Medicinal Use of Testosterone and Related Steroids Revisited." *Molecules*, vol. 26, no. 4, Feb. 2021, p. 1032. PubMed Central, <https://doi.org/10.3390/molecules26041032>.
- Te Kahu Raunui. "The Truth about Steroids." Sport Integrity Commission, 27 Jan. 2023, <https://sportintegrity.nz/news/steroids>.
- Thevis, Mario. "The 41st Manfred Donike Workshop on Doping Analysis." *Drug Testing and Analysis*, vol. 15, no. 11–12, Nov. 2023, pp. 1310–11. DOI.org (Crossref), <https://doi.org/10.1002/dta.3601>.
- Trenton, Adam J., and Glenn W. Currier. "Behavioural Manifestations of Anabolic Steroid Use." *CNS Drugs*, vol. 19, no. 7, 2005, pp. 571–95. DOI.org (Crossref), <https://doi.org/10.2165/00023210-200519070-00002>.
- Trout, Graham J., and Rymantas Kazlauskas. "Sports Drug Testing ? An Analyst's Perspective." *Chemical Society Reviews*, vol. 33, no. 1, 2004, p. 1. DOI.org (Crossref), <https://doi.org/10.1039/b201476a>.
- Turner, Diane. "GC-MS Principle, Instrument and Analyses and GC-MS/MS." *Analysis & Separations from Technology Networks*, 10 June 2022, <http://www.technologynetworks.com/analysis/articles/gc-ms-principle-instrument-and-analyses-and-gc-msms-362513>.
- Tyagi, Ritu, et al. "Metabolomics Techniques: A Brief Update." *Epigenetics and Metabolomics*, Elsevier, 2021, pp. 1–29. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-323-85652-2.00007-5>.
- UCLA. "Gas Chromatography Theory." UCLA CHEM EDU, 1 Apr. 2016, [https://www.chem.ucla.edu/~bacher/General/30BL/gc/theory.html#:~:text=The%20mobile%20phase%20\(%3Dcarrier,phase%20\(%3Dhigh%20boiling%20polymer\)](https://www.chem.ucla.edu/~bacher/General/30BL/gc/theory.html#:~:text=The%20mobile%20phase%20(%3Dcarrier,phase%20(%3Dhigh%20boiling%20polymer)).
- Urine vs Saliva Drug Testing: Pros and Cons - Australia Drug Testing. <https://www.australiadrugtesting.com/urine-vs-saliva-drug-testing/>. Accessed 23 Aug. 2024.
- Van Renterghem, Pieter, et al. "Development of a GC/C/IRMS Method – Confirmation of a Novel Steroid Profiling Approach in Doping Control." *Steroids*, vol. 77, no. 11, Sept. 2012, pp. 1050–60. DOI.org (Crossref), <https://doi.org/10.1016/j.steroids.2012.05.009>.
- Varillas-Delgado, David, et al. "Genetics and Sports Performance: The Present and Future in the Identification of Talent for Sports Based on DNA Testing." *European Journal of Applied Physiology*, vol. 122, no. 8, Aug. 2022, pp. 1811–30. DOI.org (Crossref), <https://doi.org/10.1007/s00421-022-04945-z>.
- WADA. "Anti-Doping Process." WADA, <https://www.wada-ama.org/en/athletes-support-personnel/anti-doping-process>. Accessed 25 July 2024.

The Nanjing Massacre's Effect on Contemporary Chinese Views of Japan By Jiayi Li

China and Japan are two powerful countries that fought tirelessly against each other to win, and both suffered significant losses during World War II. One event that is remembered to this day is an atrocity known as the Nanjing Massacre, or the Rape of Nanking, and it plays a significant part in shaping Chinese opinions of Japan in modern times. This essay goes over the massacre's key events and how it has been handled by historians. Finally, it will look at how the Nanjing Massacre is mentioned in contemporary Chinese sources to better understand how it influences Chinese perceptions of Japan today. To give a brief overview, this war was led by Chiang Kai-shek from 1887 to 1975, a Chinese military leader during that time, with the help of Tang Shengzhi, a Chinese warlord, under the rule of the Republic of China. Meanwhile, Japan was led by the Japanese Imperial Army. Not only did the torture that occurred during the Nanjing Massacre shape the way that the Chinese viewed the Japanese, but some events that took place afterwards also had a significant impact on their relationships. This proves how the massacre affected not only the people and society of its time but also contemporary and modern times.

This paper will first conduct a literature review. Many novels, newspapers, stories, and journalists have been published about this event. They all serve different purposes, whether to retell its story or spread awareness of this event. Some books, such as *The Battle for China*, edited by Mark Peattie, Edward J. Drea, and Hans van de Ven, discuss the Sino-Japanese component of the Second World War. There have also been many bestsellers, such as *The Good Man of Nanking: The Diaries of John Rabe* and *The Rape of Nanking*, which discuss personal stories and diaries on the Nanjing Massacre. Many primary as well as secondary resources discuss this crucial topic. However, this paper will be different in that, based on relevant sources, it will discuss how this historical event impacts contemporary Chinese views of Japan.

What Was the Nanjing Massacre?

On December 13, 1937, Nanjing (also known as Nanking), the capital city of Nationalist China at that time, fell to the Japanese. Previously, leader Chiang Kai-shek's armies were near the Yangtze Valley; they held their defenses long before being defeated (Chang 16)

Matsui Iwane, a Japanese general in the Imperial Japanese Army, ordered the massacre of Nanjing. Matsui Iwane's main helper, Tani Misato, was a lieutenant general during the killings and was involved and committed many of the terrible crimes against the Chinese in Nanjing (Ibid 39)

While Americans may date World War II back to Pearl Harbor, which occurred on December 7, 1941, and many Europeans date it from September 1, 1939, when the Nazis invaded Poland, the Chinese date it back to when the Japanese took their first step towards their domination over the continent on September 18, 1931, by invading Manchuria, a region from Northeast Asia previously under China's rule (Ibid 4)

Subsequently, numerous cities and regions fell, such as Shanghai. In 1937, fearing the loss of his military troops, Chiang removed all of them from Nanjing and fled, leaving only

untrained groups behind to defend. Nevertheless, evacuation was forbidden to citizens at this time. This event led to Nanjing's collapse and eventually the Nanjing Massacre, which occurred from December 13, 1937, to January 1938.

When the Japanese military finally took over Nanjing, their cruelty to the Chinese there was unmatched. Japanese soldiers even experimented with different ways of killing. For instance, they would often place Chinese civilians in groups of ten, bind their hands, spray them with gasoline, and set them alight, thus incinerating them.

Arguably, the worst form of torture that the Japanese used was rape. The Japanese usually targeted young or pregnant women. Still, men were included in this form of torture too as they were sometimes made to perform repulsive acts for the entertainment of the Japanese soldiers. Further, if Chinese men were caught trying to protect their loved ones from being abused, they would be killed. The torture that the Chinese endured resulted in very high casualties, with up to about 260,000 civilians murdered at the hands of the Japanese. However, some historians even claim that there had been over 350,000 casualties by the time the massacre concluded. However, this is a controversial topic, and up to this day, the Chinese and Japanese have been debating over which number—whether 260,000 or 350,000—is a more accurate death toll (Ibid 9)

At the same time, fortunately, some took steps to shelter civilians and save them from harm. Chinese citizens—along with Germans, Danish, and Russians—established the Nanking Safety Zone. Additionally, Father Jacquinet de Bessage, a French priest, established a shelter in Shanghai that took in 450,000 refugees fleeing their destroyed homes (Rabe 178). Inspired by Father de Bessage's work, W. Plumber Mills, a Presbyterian missionary, assembled a similar support group for Nanjing, making a zone within a region slightly east of the city's center into a safety zone.

After all the destruction, however, there was a significant toll on China's economy and society. Nanjing was one of the critical regions of the Chinese Republic, so the deaths led to a vast shortage of workers. This decreased trade and the number of merchants, reducing the money flow during this time, and made it difficult to repair damaged infrastructure as well as the city of Nanjing itself, much of which was rubble. It would take China years to recover.

Survival Stories

Even after the war, you can feel the effects that the massacre had on Nanjing. Du Zhongyuan, one of the best Chinese wartime journalists, wrote in November 1946, "I'd stayed three nights in Nanjing, and every night, the enemy's aircraft had invariably lit up the sky. ... When [the planes] came, a warning siren wailed. ... In the night, it was fiery all around ... the brilliance was unbelievable. I had the strong emotions I get seeing the lights at New Year (Mitter 134)." With the sirens wailing and everything in a panic, the Nanjing Massacre ravaged the city. This example shows how the bombing both immediately affected those in the city and permanently changed them mentally and physically. Li Xiuying, another survivor, claimed, "On December 19, a calamity fell upon us. ... In the morning, six Japanese soldiers came down to the basement with rifles in their hands. They took away ten young women. I was one of them. At

that time, I thought even if I had to die, I wouldn't be humiliated. I was determined to resist by dying. I hit my own head against the wall till my forehead started bleeding. I fainted and fell to the ground. When I gradually came to, the Japanese devils had left (Yang 199)." Everything the Chinese faced had a lasting impact. Li Xiuying's trauma, for instance, left permanent scars on her body and mind.

Another form of torture the Chinese had to face was watching their neighbors and friends suffer. The *North-China Herald* claimed, "The whole outlook has been ruined by frequent murder, wholesale and semi-regular looting, and uncontrolled disturbance of private homes, including offenses against the security of women ... Anyone who was caught in streets or alleys after dusk by roving patrols was likely to be killed on the spot ... The terror is indescribable (Mitter 152)." The Chinese not only suffered themselves, but they also had to watch their loved ones get hurt, causing them to always live in fear of losing their loved ones. These are only a few of the many stories that occurred during the Nanjing Massacre. While there are many Chinese sources recounting the cruelty of the massacre, however, the Japanese told a different story.

The Japanese Reaction

Ending on February 2, 1938, the Nanjing Massacre had casualties beyond comprehension. After the war ended, though, the Japanese tried to erase it from their history, preventing scholars from knowing about the Nanjing Massacre and failing to include it in their textbooks. Japan tried to cover up evidence of Nanjing, leading to some important facts being lost. To appear innocent, Japanese soldiers claimed that most victims in Nanjing were soldiers, not citizens. As a result of this erasure, few voices provided support for the rapes and the crimes committed against citizens (Yang 199). Those who were brave enough to open up about the Nanjing Massacre in Japan eventually were usually shut down and received harsh punishments, so Japanese people rarely discussed Nanjing. It was not until 1987 that the first soldier, Azuma Shiro, opened about his crimes during the Nanjing Massacre, setting an example for others who wanted to do the same. Apologizing for his crimes in public, however, led to many threats being made against him (Chang 213).

Eventually, though, the Japanese recognized the Nanjing Massacre and provided many support groups for it. For example, in the eighties, Matsuoka Tamaki investigated the Nanjing Massacre and its causes by interviewing Japanese soldiers and victims of the Nanjing Massacre alike and surveying the results. This is only one example of the Japanese taking action to make the public aware of the Nanjing Massacre (Kirk 9898). In addition to reporting, the Japanese and Chinese collaborated on newspapers. Another famous example is the *Nanking Kung Pao*, printed in 1997, which Chinese editors put together but also included information from the Japanese (Yang 203).

How The Chinese Handled This Massacre

While the Japanese only reluctantly recognized this history, the Chinese honored the Nanjing massacre and gave its victims respect a bit earlier on. From the 1980s to the present, the Chinese worked tirelessly to preserve the memory of the Nanjing massacre—starting with education. The Chinese allowed the information about the Nanjing Massacre to be widespread in China, which was the complete opposite of what the Japanese were trying to do. Scholars and citizens were informed and knew of this event, and the government built many temples and architectural structures to commemorate it, such as the Nanjing Memorial Hall, which opened in 1985 and contains objects from the atrocity. The Japanese Army invaders also contributed to creating this memorial, and it serves as a symbol of understanding and respect from both sides in this war. Unlike many other museums built during the massacre, this tourist destination solely focuses on Nanjing's people, making it unique from others (Chang 1)

In addition, many Chinese writers have tackled this difficult topic. Iris Chang, a *New York Times* bestseller for the 1997 book *The Rape of Nanking*, is one of the many great writers who recognized the importance of the Nanjing Massacre. She was the person who helped break the “six-decade-long international silence on the subject (Chang 12).” Chang worked incessantly in Nanjing, interviewing survivors and locating images and documents while agonizing over why the story was not widely known outside China. By the time she left Nanjing, Duan says, Chang was weak but even more committed to telling the story. Even though she was in poor health, she was still dedicated to spending and sharing the Nanjing survivors’ stories (Chang 12). Her book both educated Westerners and helped many Asians better understand this story. Iris Chang captured readers’ attention with pictures and the stories of the terror, evoking strong emotions.

While Iris Chang helped the outside world become aware of this massacre, she also aided the survivors. Due to her fearlessness and perseverance in writing this book, she inspired many victims of the Nanjing Massacre to speak up and share their stories. Iris Chang then recorded and shared their stories, giving them a voice and allowing them to express themselves to the world. One victim named Ni “would shake with fear trying to speak of the Japanese invasion and the slaughter of her family (Chang 1).” This shows how deeply the Nanjing Massacre affected them and how long the scars lasted. However, because of Iris Chang’s work, she inspired people like Ni to open up and share their stories and horrors.

Last but not least, Chang not only made the public more aware of the Nanjing Massacre, but she also inspired more scholars to research it. One example would be historian Sun Zhaiwei, who is a professor at the Jiangsu Academy of Social Sciences. He “was one of the first people to meet Chang in Nanjing. Sun noted that Chang uncovered the historically invaluable 2,000-page diaries of German John Rabe, a Nazi who saved tens of thousands of Chinese from certain slaughter by creating a Nanjing safety zone marshaled by the city's then-few expatriates (Chang 2).” Scholars from all over the world are now interested and willing to explore the topic.

When Iris Chang died on November 9, 2004, she was rewarded with a statue at The Memorial Hall in China (China Plus 11) Iris Chang was so monumental in memorizing the Nanjing Massacre that they honored her with a statue to allow her impact to continue impacting

people even after her death. Today, Iris Chang is remembered as one of the most influential people on the topic of the Nanjing Massacre.

The Influence of the Nanjing Massacre on Contemporary Chinese Views of Japan

Summation

The Nanjing Massacre affected contemporary times by fueling warfare between the Chinese and Japanese. Naturally, due to all the terrors that the Japanese inflicted upon the Chinese at Nanjing, the Chinese often acted aloof towards the Japanese on some occasions involving political topics, setting prejudice against each other. While the Japanese were initially cautious at approaching this piece of history, they eventually worked it out and accepted more of this piece of history. From preventing its appearance in many Japanese history textbooks to threatening the Japanese people who tried to speak up, Japan tried in different ways to belittle this momentous massacre. The Chinese tried their best to preserve this massacre and made sure that all were educated about this event. They showed this by building many temples and organizations to honor this event, such as the Nanjing Massacre Memorial. This often caused arguments and aloof relationships between the two countries as they both blame each other for this massacre, and they both think that each other is at fault. Due to all the issues between these two countries, the inhabitants of these habitats are also often distant and cool towards each other, adding fuel to their already unsettled relationship. Both China and Japan have different views on this event, and both countries believe they are right in their own way. With more than 200,000 tortured, the Nanjing Massacre was indeed an atrocity. Overall, the Nanjing Massacre deserves to be remembered and respected, which is what the two territories eventually came to realize.

Works Cited

1. Iris Chang, *The Rape of Nanking* (New York: Basic Books, 2011), 3-16.
2. *Ibid.*, 37-39.
3. *Ibid.*, 3-4.
4. *Ibid.*, 5-9.
5. John Rabe, *The Good Man of Nanking* (New York, Random House Inc., 1998) 157-178.
6. Rana Mitter, *China's War with Japan* (London, Penguin Books, 2014) 133-134.
7. Daqing Yang, *The Nanjing Atrocity* (Cambridge: Harvard University Asia Center, 2012), 199.
8. Rana Mitter, *China's War with Japan*, 152.
9. Daqing Yang, *The Nanjing Atrocity* (Cambridge: Harvard University Asia Center, 2012), 199.
10. Iris Chang, *The Rape of Nanking*, 213.
11. "Understanding Line in Art," Stanford University, <https://web.stanford.edu/~kcook/lineA.html> (accessed May 4, 2024).
12. Denton Kirk, "Exhibiting the Past: China's Nanjing Massacre Memorial Museum, Mass Violence & Résistance," *The Asia-Pacific Journal*, 1961-9898 (2014): <https://www.sciencespo.fr/mass-violence-war-massacre-resistance/en/document/exhibiting-past-chinas-nanjing-massacre-memorial-museum.html> (accessed May 4, 2024).
13. Daqing Yang, *The Nanjing Atrocity* (Cambridge: Harvard University Asia Center, 2012), 203.
14. Denton Kirk, "Exhibiting the Past: China's Nanjing Massacre Memorial Museum, Mass Violence & Résistance," *The Asia-Pacific Journal*, 1961-9898 (2014): <https://www.sciencespo.fr/mass-violence-war-massacre-resistance/en/document/exhibiting-past-chinas-nanjing-massacre-memorial-museum.html> (accessed May 4, 2024).
15. "Iris Chang, Who Chronicled Rape of Nanking, Dies at 36" *New York Times*, <https://www.nytimes.com/2004/11/12/arts/iris-chang-who-chronicled-rape-of-nanking-dies-at-36.html> (accessed May 2, 2024).
16. Iris Chang, *The Rape of Nanking*, 184-287.
17. Plus, China, "Memorial park built to honor Chinese-American writer" *China Plus*, 2019-11-11, <https://chinaplus.cri.cn/america/china-us/379390>

Making of an Athlete: Tissue-Specific Analysis of Genetic Regulators of Athletic Performance By Sri Riddhi Guruzu

Abstract

Unquestionably, researchers agree that genetic makeup is crucial in determining athletic performance abilities. Previous research has evaluated various populations through genetic analysis and revealed patterns of inheritance linking specific genetic loci to higher performance. This research explored the candidate genes that allow enhanced athletic performance across various body systems, including the muscular, nervous, and circulatory systems. An evaluation of the most studied genetic hits from each body system, including genes such as ACE, BDNF, and EPOR, was explored. The mechanism of action of each genetic loci and the variation/mutation that impacts athletic performance were examined. The findings suggested that ACE, BDNF, and EPOR strongly influence mechanisms within the muscular, nervous, and circulatory systems, respectively, enhancing athletic performance across different sports disciplines. This review also explores biological mechanisms, steroids, doping, neuro-modifications, and CRISPR, that can modify key biological substrates that underlie athletic performance in individuals. The current limitations of these biological mechanisms were addressed. Beyond genetics, it is generally understood that athletic performance success is often unpredictable due to environmental factors that affect physiological, psychological, and motor characteristics. This, combined with the ethical considerations surrounding the modification of key biological substrates involved in athletic performance highlights the dynamic landscape surrounding this area of research and the translational impact of genetics research in the field of human performance.

Introduction

Athletic performance is a multifactorial characteristic that is influenced by both nature and nurture. With respect to the former, studies show athletic performance is a complex trait and highly polygenic across multiple organ systems. To determine athletic performance abilities, organ systems are evaluated by morphology, endurance and strength capacity, metabolic activity, and other markers of organ functional and physiological capabilities. Analysis of these phenotypes allows for genetic mapping of key traits to their genetic underpinnings. Indeed, when analysis of phenotypes is paired with genetics, one is well positioned to understand key biological factors underlying athletic abilities. Genetics is crucial in determining an athlete's performance and approximately 66% of athletic performance is based on heritability(De Moor et al.). Having a favorable genetic profile can allow athletes to have favorable training, recovery, and performance outcomes that are associated with higher success in athletic competition.

While having a favorable genetic profile is not the only contributor to an athlete's success, it still remains a critical factor. Previous research has evaluated the factors that affect an athlete's success, and have found that a favorable genetic profile can not be the sole reason for competitive advantages(Maharam et al.; Gavin; Trecroci et al.). A successful athlete might have

certain genes that enable them to succeed in a particular sport, but, without years of practice and repeated training, no athlete has become successful. Moreover, the environment, nutrition, and psychological factors affect the development of a successful athlete.

Indeed, while both nature and nurture contribute to enhancing athletic abilities, this review explores nature through the genetic analysis of key organ systems important for athletic performance. In doing so, I provide mechanistic insights into how certain genes within key organ systems are involved in athletic performance.

Gene	Full Name	Locus	Polymorphism	Endurance-Related Allele	Mechanism of action	References	
						Studies with Positive Results	Studies with Negative or Controversial Results
ACE	Angiotensin I converting enzyme	17q23.3	Alu I/D (rs4343 A/G or rs4341 C/G)	I (A or C)	It encodes an enzyme involved in blood pressure regulation and electrolyte balance(ACE Angiotensin I Converting Enzyme [Homo Sapiens (Human)] - Gene - NCBI).	(Montgomery et al.; Gayagay et al.; Nazarov et al.; Myerson et al.; Jelaković et al.; I. I. Ahmetov, Popov, et al.; Alvarez et al.; Collins et al.; Lucia et al.; The Angiotensin Converting Enzyme I/D Polymorphism in Long Distance Runners - ProQuest; Scanavini et al.; Turgut et al.; Tsianos et al.; Cieszczyk et al.; Min et al.; Shenoy; Znazen et al.)	(Lucia et al.; Ash et al.; Tobina et al.; Ildus I. Ahmetov et al.; The ACE I/D Polymorphism in Elite Greek Track and Field Athletes Request PDF; R. A. Scott et al.; Rankinen et al.; Taylor: Elite Athletes and the Gene for Angiotensin-Conve... - Google Scholar; The Association between Ace Gene Variation and Aerobic Capacity in Winter Endurance Disciplines - PubMed; Ginevičienė, Pranculis, et al.; Muniesa et al.; Ioannis D. Papadimitriou, Lucia, et al.; Varillas-Delgado et al.)

Gene	Full Name	Locus	Polymorphism	Endurance-Related Allele	Mechanism of action	References	
						Studies with Positive Results	Studies with Negative or Controversial Results
ACTN3	Actinin α 3	11q13.1	rs1815739 C/T	T	It encodes a member of the alpha-actin which is primarily expressed in skeletal muscle and functions as a structural component of the sarcomeric Z line(ACTN3 Actinin Alpha 3 [Homo Sapiens (Human)] - Gene - NCBI).	(N. Yang, MacArthur, et al.; Shang et al.; Gasser et al.; Orysiak et al.)	(R. Yang et al.; G. Wang et al.; Mikami et al.; Grealy et al.; Lucia et al.; N. Yang, Macarthur, et al.; C. J. Saunders et al.; Papparini et al.; I. D. Papadimitriou et al.; Niemi and Majamaa; Tsianos: Associations of Polymorphisms of Eight Muscle-or... - Google Scholar; Ginevičienė: Relating Fitness Phenotypes to Genotypes... - Google Scholar; F. E. Döring et al.; I. I. Ahmetov, Druzhevskaya, et al.; Ioannis D. Papadimitriou, Lockey, et al.)
ADRB2	Adrenoc eptor β 2	5q31-q32	rs1042713 G/A	A	It encodes beta-2-adrenergic receptor which is associated with the class C L-type calcium channel Ca(V)1.2(ADRB 2 Adrenoceptor Beta 2 [Homo Sapiens (Human)] - Gene - NCBI).	(Tsianos: Associations of Polymorphisms of Eight Muscle-or... - Google Scholar; Wolfarth et al.; Wagoner et al.)	(Santiago et al.; Sawczuk, Maciejewska-Karłowska, et al.)
AMPD1	Adenosine monophosphate deaminase 1	1p13	rs17602729 C/T	C	It catalyzes the deamination of AMP to IMP in skeletal muscle and plays an important role in the purine nucleotide cycle(AMPD1 Adenosine Monophosphate Deaminase 1 [Homo Sapiens (Human)] - Gene - NCBI).	(Varillas-Delgado et al.; Rubio et al.; Thomaes et al.; Rico-Sanz et al.; Ciężczyk et al.)	(Ginevičienė, Jakaitienė, et al.)

Gene	Full Name	Locus	Polymorphism	Endurance-Related Allele	Mechanism of action	References	
						Studies with Positive Results	Studies with Negative or Controversial Results
BDKR B2	Bradykinin receptor B2	14q32.1-q32.2	9/-9 (exon 1)	-9	It encodes a receptor for bradykinin that stimulates a phosphatidylinositol-calcium second messenger system(BDKRB2 Bradykinin Receptor B2 [Homo Sapiens (Human)] - Gene - NCBI).	(Williams et al.; Colleen J. Saunders et al.)	(Varillas-Delgado et al.; Sawczuk, Timshina, et al.; Grenda: Bdkrb2 Gene-9/+ 9 Polymorphism and Swimming... - Google Scholar; Zmijewski, Grenda, et al.)
HFE	Homeostatic iron regulator	6p21.3	rs1799945 C/G	G	It encodes a membrane protein that associates with beta2-microglobulin (beta2M)(HFE Homeostatic Iron Regulator [Homo Sapiens (Human)] - Gene - NCBI).	(Semenova, Miyamoto-Mikami, et al.; Hermine et al.; Chicharro et al.; Deugnier et al.; Varillas-Delgado et al.)	
HIF1 A	Hypoxia inducible factor 1 subunit alpha	14q23.2	rs11549465 C/T	C	It encodes the alpha subunit of HIF-1 which functions as a master regulator of cellular and systemic homeostatic response to hypoxia(HIF1A Hypoxia Inducible Factor 1 Subunit Alpha [Homo Sapiens (Human)] - Gene - NCBI).	(F. Döring et al.; Prior et al.)	(Ildus I. Ahmetov et al.; Lack of Association between the GNB3 Rs5443, HIF1A Rs11549465 Polymorphisms, Physiological and Functional Characteristics - Bosnyák - 2020 - Annals of Human Genetics - Wiley Online Library)

Gene	Full Name	Locus	Polymorphism	Endurance-Related Allele	Mechanism of action	References	
						Studies with Positive Results	Studies with Negative or Controversial Results
NOS3	Nitric oxide synthase 3	7q36	rs2070744 T/C	T	It acts as a biological mediator in several processes, including neurotransmission and antimicrobial and antitumoral activities(NOS3 Nitric Oxide Synthase 3 [Homo Sapiens (Human)] - Gene - NCBI).	(Varillas-Delgado et al.; Zmijewski, Cięszczyk, et al.; Drozdowska: Allelic Polymorphism of Endothelial NO-Syntha... - Google Scholar)	(Gómez-Gallego et al.)
PPAR A	Peroxisome proliferator activated receptor α	22q13.31	rs4253778 G/C	G	It encodes the subtype PPAR-alpha which is a nuclear transcription factor((PPARA Peroxisome Proliferator Activated Receptor Alpha [Homo Sapiens (Human)] - Gene - NCBI).	(Tural et al.; Maciejewska, Sawczuk, and Cięszczyk; Akhmetov: Association of Regulatory Genes Polymorphisms... - Google Scholar; Ildus I Ahmetov et al.)	
PPAR GC1A	Peroxisome proliferative activated receptor, γ , coactivator 1 α	4p15.1	rs8192678 G/A	G	It encodes a transcriptional coactivator protein that regulates the genes involved in energy metabolism(PPARGC1A PPARG Coactivator 1 Alpha [Homo Sapiens (Human)] - Gene - NCBI).	(Maciejewska, Sawczuk, Cieszczyk, et al.; Akhmetov: Association of Regulatory Genes Polymorphisms... - Google Scholar; Lucia: PPARGC1A Genotype (Gly482Ser) Predicts Exceptional... - Google Scholar; R. Yang et al.)	(Hall et al.; He et al.; Mitochondrial DNA Variation Is Associated with Elite Athletic Status in the Polish Population - Maruszak - 2014 - Scandinavian Journal of Medicine & Science in Sports - Wiley Online Library)

Gene	Full Name	Locus	Polymorphism	Endurance-Related Allele	Mechanism of action	References	
						Studies with Positive Results	Studies with Negative or Controversial Results
BDNF	Brain derived neurotrophic factor	11p14.1	rs10501089 G/A	A	Plays an important role in neuronal survival and growth, serves as a neurotransmitter modulator, and participates in neuronal plasticity(Bathina and Das).	(Guilherme et al.)	
EPOR	Erythropoietin receptor	19p13.2	N/A	N/A	The receptor activates Jak2 tyrosine kinase which activates different intracellular pathways(PubChem, EPOR - Erythropoietin Receptor (Human)).	(Truncated Erythropoietin Receptor Causes Dominantly Inherited Benign Human Erythrocytosis. - PMC)	

Table 1: An overview of most studied genes that influence endurance performance, including their locus, polymorphism, allele, and mechanism of action. Table recreated from Semenova et al., 2023(Semenova, Hall, et al.)

Overview of Muscle Biology

The muscular system is one of the most vital systems in the human body. It allows for the movement of organisms but also aids in maintaining posture, ensuring joint stability, and producing heat(Introduction to the Muscular System | SEER Training). The human body is composed mainly of 3 types of muscles: skeletal muscles, smooth muscles, and cardiac muscles, and each serves a unique purpose.

Cardiac muscle, also known as myocardium, is present between the pericardium and endocardium layers in the heart. It consists of intercalated discs that contract together synchronously, t-tubules in the sarcolemma for excitation-contraction coupling, and multiple mitochondria to generate energy(Ripa et al.). These unique specializations allow it to generate sufficient force and coordinate contractions to pump blood around the body.

Comparatively, skeletal muscles are present throughout the body. A skeletal muscle fiber is composed of multiple microfibrils containing actin (thin filaments), myosin (thick filaments), and support proteins that allow movement and sustain body posture and position (McCuller et al.).

Lastly, smooth muscles are present in the gastrointestinal, reproductive, urinary, vascular, and respiratory systems. They consist of a calcium-containing sarcoplasmic reticulum, which aids in sustaining contraction, and actin and myosin, which act as the main proteins involved in muscle contraction(Hafen et al.).

The major difference between these muscle types is in their appearance and mechanisms. Cardiac and Skeletal muscles have a striated appearance, while smooth muscles have a non-striated appearance. Mechanically speaking, cardiac and smooth muscles contract involuntarily, unlike skeletal muscles, which contract voluntarily(Muscle Types | SEER Training). Moreover, smooth muscle contracts slowly and rhythmically while cardiac muscle contracts strongly and rhythmically(Muscle Types | SEER Training). An underlying reason for these different features differences is differences in gene expression and development.

Relevance of slow twitch muscle fibers to athletic training

All skeletal muscles are composed of specialized cells known as muscle fibers. There are primarily two types of skeletal muscle fibers: slow-twitch muscle fibers, and fast-twitch muscle fibers. Slow-twitch fibers are the smallest fiber type and have a low glycogen content. On the other hand, fast-twitch muscle fibers are the largest fiber type due to their high density of actin and myosin proteins and have a higher glycogen content than slow-twitch muscle fibers(W. Scott et al.). While slow-twitch fibers are more resistant to fatigue and contract for longer periods, fast-twitch fibers sustain short anaerobic bursts of activity and fatigue easier. In other words, slow-twitch muscle fibers are associated with long-distance running while fast-twitch muscle fibers are associated with sprinting(Ji).

Multiple studies in the past have indicated that endurance athletes have enhanced slow-twitch muscle fibers(Harber et al.; Harber and Trappe; Semenova, Zempo, et al.). In one paper by Harber et al., researchers examined five male runners (age = 20 ± 1 yr; wt = 65 ± 4 kg; ht = 178 ± 3 cm) of a men's varsity cross-country team and collected muscle biopsies to isolate and study the muscle fibers in different training periods. They found an increase in levels of slow-twitch and fast-twitch fibers from the period of high-volume base training (T1), to the 8-week high-intensity training (T2), and to the 4-week reduced training phase (T3). Similarly, M. Harber & Trappe divided the sixteen subjects into 2 groups- trained competitive male runners and recreationally active female and male runners- and collected muscle biopsies to isolate and analyze the myofibers. They found higher slow-twitch muscle fiber composition in the trained competitive runners than the recreationally active runners. More recently, Semenova et al. involved 219 elite Russian athletes, 114 Japanese athletes, and 287 Finnish study athletes, and each cohort was divided into power athletes and endurance athletes. GWAS was used to evaluate the muscle fiber compositions in each population. It concluded that endurance athletes have a higher proportion of slow-twitch muscle fibers than power athletes. This study has a large sample size ranging over a diverse population and provides robustness by conducting an extensive data analysis.

Studies on the Genetics of Muscle Biology

Recent studies in 2022 and 2023 by Semenova et al. (Semenova, Zempo, et al.; Semenova, Hall, et al.) set out to study the single nucleotide polymorphisms underlying the inter-individual differences in muscle fiber types and the genetic markers (DNA polymorphisms) associated with athlete status, respectively. The studies used genome-wide association studies (GWAS) to determine the genetic markers associated with differences in muscle types and genes associated with endurance athlete status. From these studies, the top 5 gene hits are vascular endothelium growth factor receptor 2 (VEGFR2), peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PPARGC1A), peroxisome-induced alpha receptor (PPARA), alpha-actinin 3 (ACTN3), and angiotensin-converting enzyme (ACE). VEGFR-2 is the major receptor of VEGF mediates a variety of signaling transduction, biological responses, and pathological processes in angiogenesis (X. Wang et al.). It is expressed in vascular endothelial cells. PPARGC1A gene encodes a transcriptional coactivator protein that regulates the genes involved in energy metabolism (PPARGC1A PPARG Coactivator 1 Alpha [Homo Sapiens (Human)] - Gene - NCBI). It interacts with PPARgamma and is highly expressed on the surface of the intestinal epithelium (D'Errico et al.). PPARAs are nuclear receptors that correlate with high mitochondrial and peroxisomal β -oxidation activities (Desvergne and Wahli). They are expressed in the liver, skeletal muscle, heart, and muscle. ACTN3 encodes the protein alpha-actinin-3 which plays an important role in impacting the muscle phenotype in elite athletes and its expression is restricted to fast-twitch muscle fibers (ACTN3 - an Overview | ScienceDirect Topics). Lastly, the ACE gene encodes the angiotensin-converting enzyme, which converts angiotensin I into angiotensin II.

Multiple studies suggest that ACE is a recurring genetic hit for muscle fibers in athletic performance (Zhang et al.; Kumagai et al.). In a paper by Zhang et al., researchers examined untrained healthy young volunteer subjects (31 males, 10 females, age 24 ± 3 years) by collecting skeletal muscle samples using the needle-biopsy method. They found that the ACE-I allele was associated with increased slow-twitch muscle fiber. More recently, Kumagai et al., included 211 healthy Japanese individuals (102 men and 109 women) to collect their muscle biopsies and analyze multiple polymorphisms including ACE. The researchers found that ACE I/D polymorphisms influence the muscle fiber composition in Japanese men. ACE has been a frequent genetic hit for muscle fiber composition. Hence, ACE will be examined further in subsequent sections.

ACE Biology and Impact on Performance

A polymorphism in the ACE gene strongly influences human physical performance. Specifically, an 'insertion' allele of ACE is associated with lower tissue ACE activity, and improved endurance performance (F. Ma et al.; Eider et al.). ACE catalyzes the conversion of angiotensin I into a physiologically active peptide angiotensin II and degrades active bradykinin (BK) (Wong). The enzyme plays a major role in the renin-angiotensin-aldosterone system (RAAS). The RAAS is a crucial mediator of cardiac, vascular, and renal physiology through

regulating vascular tone and salt and water homeostasis(Fountain et al.). This pathway is outlined in Figure 1 and begins with Angiotensinogen, a molecule primarily synthesized and constitutively secreted by the liver. Angiotensinogen is broken down into Angiotensin I by Renin, an enzyme expressed in the kidney. Then angiotensin I into angiotensin II by the activity of ACE. Angiotensin II is the primary mediator of blood pressure, volume regulation, and aldosterone secretion in the human body(Fountain et al.). ACE2, a surface protein expressed on different tissues and organs(Angiotensin-Converting Enzyme 2 - an Overview | ScienceDirect Topics), breaks down Angiotensin II into Angiotensin-(1-7). There are two types of angiotensin II receptors: Angiotensin type 1 receptor (AT1 receptor) and type 2 receptor (AT2 receptor)(Singh and Karnik). The AGTR2 gene encodes the AT2 receptor. Previous research demonstrated that the AGTR2 gene C allele is associated with an increased proportion of slow-twitch muscle fibers, and the A allele is associated with a higher percentage of fast-twitch fibers(Mustafina et al.). This pathway is outlined in Figure 1.

The ACE gene insertion/deletion (I/D) polymorphism has been associated with variation in the ACE plasma level. These alleles account for approximately half (47%) of the observed variance in ACE levels(Ahmad Yusof and Che Muhamed; Thakur et al.). The insertion allele decreases the ACE plasma level, reducing the skeletal muscle vasoconstriction and thus increasing oxygenated blood supply to working muscles(Ahmad Yusof and Che Muhamed). On the other hand, the deletion allele of ACE increases the ACE plasma level, is associated with higher angiotensin II levels, and results in skeletal muscle hypertrophy(Ahmad Yusof and Che Muhamed). Research has also shown that the ACE-I allele is associated with increased slow-twitch fibers(Zhang et al.; Vaughan et al.) and the ACE-D allele is associated with increased fast-twitch fibers(Ioannis D. Papadimitriou, Lucia, et al.). These variations in ACE impact the muscle fiber composition and muscle oxygenation, resulting in improved athletic performance.

An increase in the slow-twitch muscle fiber composition allows elite athletes to stay more resistant to fatigue and contract for longer periods. This would, in theory, improve the endurance performance of the athletes and positively impact their athletic ability. Indeed, previous studies suggest that athletes in endurance-related sports had an excess of the ACE I allele than in other sports(Montgomery et al.; Myerson et al.; The Angiotensin Converting Enzyme I/D Polymorphism in Long Distance Runners - ProQuest). A study by Hruskovicová et al., examined marathon runners, half-marathon runners, and inline skaters to study the association of ACE I/D polymorphism with enhanced endurance performance. They concluded that there was an excess of the ACE I allele in long-distance runners. Similar studies are cited in Table 1 and they outline ACE and its impact on athletic performance.

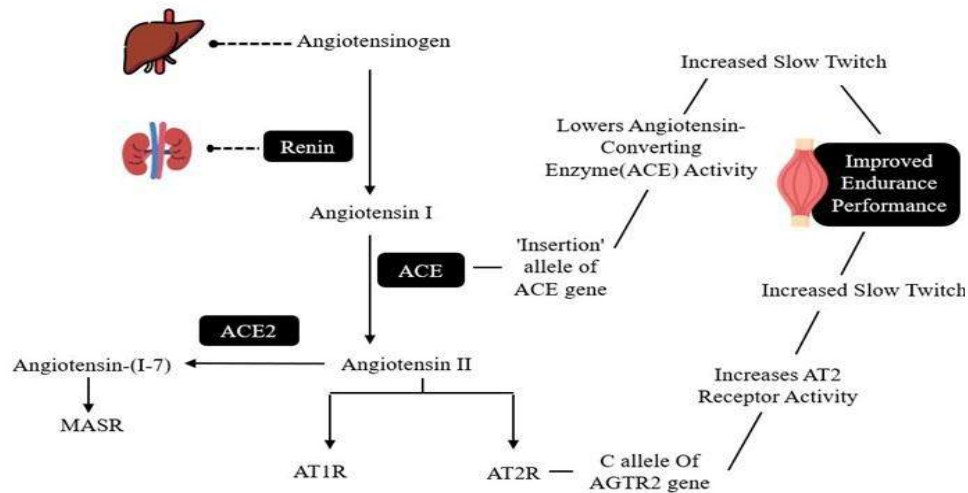


Fig 1: Schematic representation of ACE biochemical pathway.

Overview of the Nervous System

The nervous system is a critical component of the human body that performs many essential functions, including emotive and cognitive processing, regulation of key physiological activities, and planning and shaping of motor movements and skills. The latter, planning and shaping of motor movements and skills, is essential for athletic performance and training and is described in further detail in subsequent sections. The structures that allow the nervous system to participate in these complex phenomena include the central (brain and spinal cord) and peripheral (nerves) nervous systems. The peripheral nerves innervate key organs and muscle systems that allow for relaying information to the spinal cord and the brain from the target tissues, and importantly also allow for top-down modulation of the target tissues by the brain and spinal cord. The exact mechanisms and processes will be discussed in subsequent sections, but generally reference fundamental neuroplasticity mechanisms that underlie skill acquisition. In brief, both the central and peripheral nervous systems are capable of facilitating such processes through their constitutive elements, neurons, and glial cells.

Neurons play a critical role in communication within the nervous system and establish coordination using a combination of sensory, motor, and interneurons. Sensory neurons are associated directly with receptors and convey messages regarding external stimuli (Sensory Neuron - an Overview | ScienceDirect Topics). They usually have long dendrites and relatively short axons to quickly send signals from the stimuli to the central nervous system. Conversely, motor neurons carry messages to control the movement through muscles (Davidovits). They usually have short dendrites and long axons to increase the surface area that can receive signals and reach the target muscle groups. Overall, the specializations within the nervous system allow humans to improve coordination, speed, and memory.

Relevance of the Nervous System for Athletic Training

Cognitive skills, including decision-making, action anticipation, and rapid information processing, are crucial for the performance of elite athletes. Researchers have studied the structural differences and predictive judgments in athletes and found that the brains of athletes are anatomically and functionally different from those of non-athletes.

A study by Paruk et al., aimed to compare the brains of those who engage in extremely high levels of endurance exercise with those who are sedentary using magnetic resonance imaging(Paruk et al.). They found an overall increase in grey and white matter volumes in athletes. Grey matter consists of neural somas while white matter consists of myelinated axons and both of them play a key role in processing information. Another study by Giesler et al. examined the correlation between cardiorespiratory fitness levels and grey and white matter volumes in young endurance athletes and non-athletes(Geisler et al.). Using diffusion magnetic resonance imaging, they found that the white matter volume was larger and the gray matter volume was smaller in the athletes than in non-athletes. Similarly, a neuroimaging study by Park et al. looked into the gray matter (GM) and white matter (WM) volumes in elite basketball players and found significantly larger WM volumes in their vermian lobules VI-VII(Park et al.). A more surprising study by Wei et al. examined the cortical thickness in diving athletes and non-athletes and found an increased cortical thickness in the athlete group(Increased Cortical Thickness in Sports Experts: A Comparison of Diving Players with the Controls | PLOS ONE). Multiple studies suggest that athletes' brains undergo structural changes with practice.

Other researchers examined the action anticipation in elite athletes from professional sports(Müller et al.; Aglioti et al.). Müller et al. studied the predictive judgments made by low-skilled, intermediate, and highly-skilled cricket batsmen on the type and length of balls bowled by bowlers. Through a combination of temporal occlusion of the display and selective occlusion, the researchers concluded that highly skilled players picked up early cues; however, the intermediate and low-skilled players lacked this capability. Aglioti et al. conducted a similar study on elite basketball athletes, expert watchers (which included 5 coaches and 5 sports journalists), and novices. They found only elite athletes showed a time-specific motor activation while observing the free shots at a basket. This fascinating research proves that athletes have faster predictive judgments.

Studies on the Genetics of Neuron Biology

It has been evident in multiple studies that genes contribute to the success of athletes in specialized sports. The previous sections explored the genes that impact muscle fiber composition in athletic performance; however, this section will explore the key genes that contribute to the central nervous system for enhanced performance. A recent study conducted by Kitazawa et al. set out to study the genetic variations implicated in the central nervous system of elite athletes(Kitazawa et al.). The study accumulated evidence from multiple studies on genes associated with enhanced performance in the central nervous system.

The top 5 gene hits are Dopamine D2 receptor (DRD2), β 1 adrenergic receptor (ADRB1), Fifth Ewing variant (FEV), Catechol-O-methyltransferase (COMT), and Brain-derived neurotrophic factor (BDNF). DRD2 encodes the D2 subtype of the dopamine receptor (DRD2 Dopamine Receptor D2 [Homo Sapiens (Human)] - Gene - NCBI), is primarily expressed in the adrenal tissue, and it has been known to improve Motor learning and performance (Jacob et al.; Noohi et al.). ADRB1 gene polymorphisms have been shown to affect the resting heart rate (ADRB1 Adrenoceptor Beta 1 [Homo Sapiens (Human)] - Gene - NCBI) and a rare mutation in this gene has regulated sleep (Shi et al.). These receptors are predominantly located in the heart. FEV leads to elevated serotonergic activity for optimal performance in Polish athletes (Peplonska et al.). The gene is exclusively expressed in neurons of the central serotonin (5-HT) system (FEV FEV Transcription Factor, ETS Family Member [Homo Sapiens (Human)] - Gene - NCBI). COMT catalyzes the transfer of a methyl group from S-adenosylmethionine to catecholamines, including the neurotransmitters dopamine, epinephrine, and norepinephrine, and it is ubiquitously expressed in the placenta (COMT Catechol-O-Methyltransferase [Homo Sapiens (Human)] - Gene - NCBI). Lastly, the BDNF gene encodes a member of the nerve growth factor family of proteins and may play a role in regulating stress responses (PubChem, BDNF - Brain Derived Neurotrophic Factor (Human)).

Multiple studies suggest that BDNF is a gene relevant to the central nervous system and athletic performance (Schor et al.; Muñoz Ospina and Cadavid-Ruiz). In a paper by Schor et al., the researchers aimed to collect the plasma BDNF before and after a judo training session (Randori) and the maximal incremental ramp test (MIRT) in athletes from the Brazilian national judo team. They found a greater increase in BDNF after Randori than MIRT, suggesting that increases in BDNF levels are specific to the type of training.

More recently, a study conducted in 2024 aimed to determine the effect of a single session of acute aerobic exercise on serum BDNF levels in college students with varying levels of physical activity. The researchers divided the samples into 3 groups: athletes (n = 20), regular fitness (n = 19), and sedentary (n = 23). They concluded that active young adults and athletes demonstrate low serum BDNF concentration baselines and sufficient sensitivity to increase BDNF concentration with a single exercise session.

BDNF Biology and Impact on Performance

In the human body, BDNF promotes the survival of peripheral sensory neurons during the development of the brain and promotes angiogenesis (Bathina and Das). In athletic performance, however, it plays a unique role. More evident research suggests that BDNF plays an important role in regulating neuroplasticity. Regulating neuroplasticity in athletes has been found to improve memory, reduce atrophy, and elevate cognitive processing (Miranda et al.). In addition, research has found that the overexpression of BDNF results in an increased proportion of fast-twitch muscle fibers. The researchers collected muscle biopsies of 508 Russian power athletes, 178 endurance athletes, and 190 controls; they found that the minor A-allele had a

significantly higher percentage of fast-twitch muscle fibers than individuals homozygous for the G-allele(Guilherme et al.). Figure 2 outlines BDNF's role in athletic performance.

When BDNF to TrkB (a neurotrophin receptor tyrosine kinase), it causes the activation of several downstream signaling cascades(Chakrapani et al.). This promotes the development of favorable neuroplastic phenomena, for instance, local protein synthesis, spine remodeling, or gene transcription(Miranda et al.). Moreover, the release of BDNF can help promote cell survival and establish smooth communication between cardiovascular and skeletal muscle activity and the central nervous system, which could impact the proportion of slow-twitch muscle fiber composition. Beyond muscle compositional changes, it is also possible given BDNF's role in neuroplasticity that the structural, functional, and behavioral outcomes observed in athletes may be associated with genetic variations in BDNF signaling.

To summarize, acute aerobic exercise in particular sports has increased BDNF levels in the brain. BDNF is crucial in promoting cell survival, developing favorable neural plasticity, reducing atrophy, and improving memory. Expression of the minor A-allele of BDNF has been associated with higher fast-twitch composition and increased hand-grip strength. Moreover, neural plasticity aids elite players from various sports, including enhancing temporal processing in tennis players(Overney et al.), changing cortical grey matter and white matter architecture in golfers(Jäncke et al.), and enhancing action anticipation and motor resonance in basketball players(Aglioti et al.). See Figure 2 for a graphical depiction of these findings.

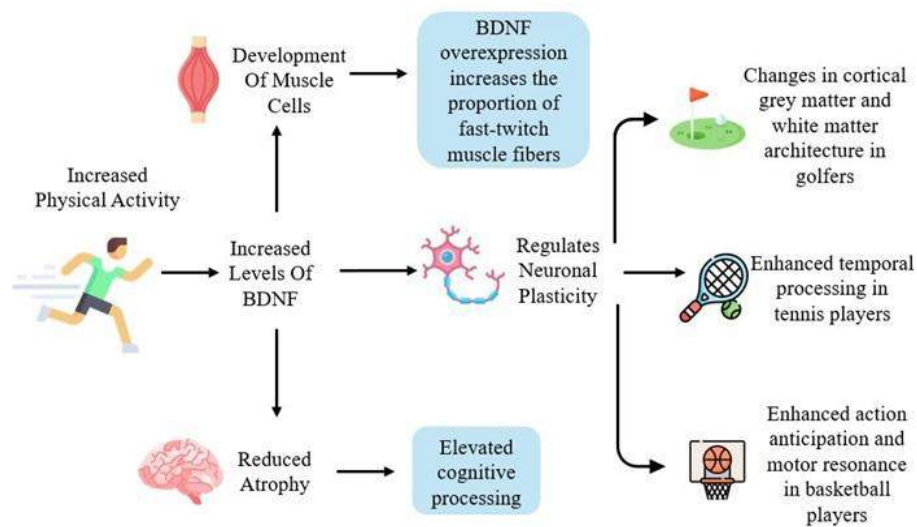


Fig 2: Schematic representation of BDNF biochemical pathway.

Overview of Blood Cell Biology

The circulatory system ensures the survival of all cells of the body by maintaining the immediate chemical environment of each cell in the body(Pittman). It consists of the heart for pumping blood and blood vessels for carrying nutrients to the entire body. The liquid portion of

blood that consists of water, proteins, electrolytes, dissolved gas, and nutrients, is known as blood plasma, and transports three types of blood cells: red blood cells, white blood cells, and platelets.

Red blood cells (RBCs), or erythrocytes, are the components of blood responsible for gas and nutrient transportation in the human body(Barbalato and Pillarisetty). RBCs transport oxygen from the lungs to the peripheral tissues to assist in metabolic processes, such as motor movements and skills. They are specialized, non-nucleated, biconcave-shaped cells that increase surface area for oxygen diffusion, and optimize their flow properties in vessels and capillaries(Normal Red Blood Cells' Shape Stabilized by Membrane's in-Plane Ordering | Scientific Reports). They contain hemoglobin, a globular protein, that consists of two α -subunits ($\alpha 1$ and $\alpha 2$) and two β -subunits ($\beta 1$ and $\beta 2$) that are structurally similar, and an iron-heme component, which allows the binding of oxygen to form an octahedral iron complex.

White blood cells (WBCs), or leukocytes, mount inflammatory and cellular responses to injury or pathogens and participate in the innate and humoral immune responses(Tigner et al.). They have an irregular shape that allows them to engulf pathogens, and they replicate quickly to help fight infectious diseases. WBCs are classified into three types: granulocytes, monocytes, and lymphocytes. Each of them is specialized for a unique function in the body. Platelets play a key role in hemostasis by serving as an immediate reparative response to injury of the vascular system designed to prevent blood loss(Williams and Sergent). They are produced in the bone marrow and have multiple specializations. The discoid-shaped cells are 1.5–3 μm in size, and contain granules that can secrete proteins required for creating a sealed mesh for blood vessel breaks. This allows them to aid in blood clotting and wound healing. These three major components of blood allow the proper functioning of the circulatory system.

Relevance of the cardiovascular system to athletic training

The cardiovascular system plays a major role in athletic performance. Previous studies indicate that elite athletes have improved blood pressure, heart rates, blood volume, hemoglobin mass, and greater posterior wall thickness than untrained individuals, enhancing their endurance performance(Maharam et al.; Heinicke et al.; Douglas and O'Toole).

Heinicke et al., examined the total hemoglobin and blood volume in 94 male elite athletes subdivided into different disciplines using the CO-rebreathing method(Heinicke et al.). The method consists of inhaling and rebreathing a bolus of CO through a spirometer for 2 min and an analysis of the increase of carboxyhemoglobin (COHb) content of capillary blood at about 7 min after inhalation of CO(Ahlgrim et al.). The researchers compared the results of the elite athletes with untrained individuals and found that the total hemoglobin and blood volume were 35 - 40% higher in the endurance groups than in untrained individuals. Likewise, another research paper evaluated the effects of physical activity on cardiac function and structure in athletes and non-athletes using guided M-mode echocardiograms and pulsed Doppler studies(Douglas and O'Toole). They found that athletes had lower heart rates, larger left ventricular cavities, and

higher ratios of early to atrial inflow velocities than non-athletes. This research brings to light the impact of exercise on cardiac health.

Studies on the Genetics of Blood Cell Biology

It is evident from previous research that genes can impact the circulatory system, allowing for improved performance in athletes. In a research paper by Malczewska-Lenczowska et al., the association between -551C/T and intron 2, +16 C/G polymorphisms in the beta hemoglobin (HBB) gene and total hemoglobin mass (tHbmass) was examined in endurance athletes (Malczewska-Lenczowska et al.). They sampled 89 young road cyclists, female (n = 39) and male (n = 50), and determined 2 polymorphisms in the HBB gene. The researchers found that the HBB gene could be related to aerobic capacity.

Furthermore, multiple studies examined the effects of mutations in the EPOR gene on the production of RBCs. A family with autosomal dominant erythrocytosis was studied, and one of the affected family members had an increased sensitivity to erythropoietin causing benign erythrocytosis (Juvonen et al.). The mechanisms of EPOR and its role in athletic performance will be evaluated in subsequent sections.

EPOR Biology and Impact on Performance

EPOR gene encodes the erythropoietin receptor which is expressed at the cell surface as a dimer in the absence of a ligand (Watowich). The erythropoietin receptor is a single-span membrane protein that plays a crucial role in the EPO/EPOR signaling pathway. In this pathway, erythropoietin (EPO) regulates the proliferation and differentiation of erythroid precursor cells in the body. When EPO binds to the erythropoietin receptor, JAK 2 kinase is activated and it stimulates other pathways including STAT5, PI3K/Akt, NF- κ B, and MAPK (C. Ma et al.). All of these pathways lead to the proliferation of RBCs in the body.

Erythrocytosis or polycythemia is the absolute or relative increase in hemoglobin (Hgb) and hematocrit (Hct) levels (Gangat et al.). It could be caused due to rare congenital mutations, respiratory diseases, cardiac disease, renal disorders, or exogenous administration of erythropoietin (Keohane et al.). As mentioned previously, a rare mutation in EPOR was found to be associated with benign erythrocytosis. The intracellular C-terminal part contains a domain exerting negative control on erythropoiesis and when a G to A transition occurs in nucleotide 6002 of the EPOR gene, a mutation occurs (Truncated Erythropoietin Receptor Causes Dominantly Inherited Benign Human Erythrocytosis. - PMC). This causes increased hemoglobin levels and oxygen intake, as outlined in Figure 3.

An excess number of red blood cells could slow blood flow, and increase the likelihood of blood clots (Byrnes and Wolberg). However, a controlled increase in hemoglobin levels allows faster oxygen transportation to cells around the body, including muscle cells. The Finnish cross-country skier Eero Mäntyranta had a mutation in the EPOR gene that increased the number of red blood cells. Increasing oxygen intake directly improves endurance performance. This is thought to contribute to his athletic success, as he won 7 medals in 4 Winter Olympics.

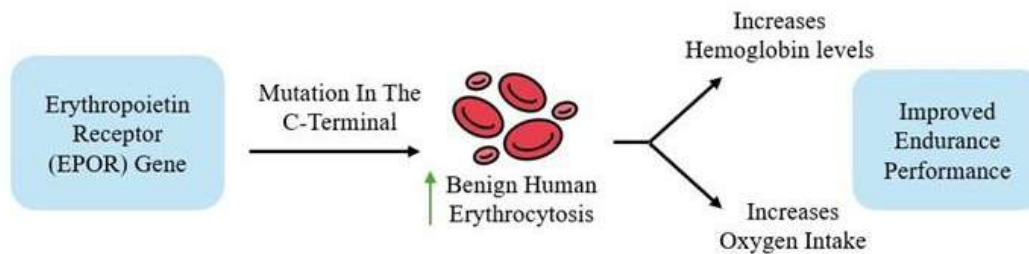


Fig 3: Schematic representation of EPOR biochemical pathway.

Discussion

Motion is a fundamental output of the human body and one that is perhaps most magnified and dissected at the level of professional athletes. Indeed, the performance of athletes is studied from diverse perspectives, from biological to data analytics, in part to understand and characterize features that can improve key outputs in a given sport or athletic competition. Additionally, characterization of the factors that underlie high athletic performance allows for evidence of training and developmental regimens that can improve junior and novice athletes' gain of critical skills and abilities for their craft. As such, this review article has explored these core concepts of features that drive high-performance athletes through the lens of key organ systems and genetic factors that regulate these systems.

As discussed earlier, the muscle system is critical for the movement of organisms but also aids in maintaining posture, ensuring joint stability, and producing heat (Introduction to the Muscular System | SEER Training). Slow-twitch fibers are more resistant to fatigue and contract for longer periods, and they are associated with endurance performance. Fast-twitch fibers sustain short anaerobic bursts of activity and fatigue more easily and are associated with strength performance. The following genetic hits influence the muscle fiber composition: vascular endothelium growth factor receptor 2 (VEGFR2), peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PPARGC1A), peroxisome-induced alpha receptor (PPARA), alpha-actinin 3 (ACTN3), and angiotensin-converting enzyme (ACE). Specifically, the “insertion allele” of ACE strongly enhances endurance performance.

Additionally, the nervous system is a critical component of the human body that performs emotive and cognitive processing, regulation of key physiological activities, and planning and shaping of motor movements/skills. Athletes' brains are anatomically and functionally different from those of non-athletes. The following genetic hits influence the variations implicated in the central nervous system of elite athletes: Dopamine D2 receptor (DRD2), β 1 adrenergic receptor (ADRB1), Fifth Ewing variant (FEV), Catechol-O-methyltransferase (COMT), and Brain-derived neurotropic factor (BDNF). Specifically, BDNF is crucial in promoting cell survival, developing favorable neural plasticity, reducing atrophy, and improving memory.

Moreover, the circulatory system ensures the survival of all cells of the body at every moment by maintaining the immediate chemical environment of each cell in the body(Pittman). Elite athletes have improved blood pressure, heart rates, blood volume, hemoglobin mass, and greater posterior wall thickness than untrained individuals. Particularly, a rare mutation in EPOR was found to be associated with benign erythrocytosis (an increase in hemoglobin levels).

Athletes with a favorable variation in ACE, BDNF, or EPOR would be at a competitive advantage when compared to other athletes without favorable variations. This was especially evident in the mutation in EPOR, which contributed to the success of a Finnish cross-country skier, Eero Mäntyranta.

While having favorable genes is extremely rare, genes are not the sole influencers of athletic performance and can be unpredictable due to environmental factors or nurture. All the benign physiological, psychological, and motor characteristics in athletes must align for athletes to have a competitive advantage.

Nevertheless, athletic performance can be modulated biologically through a diverse range of methods, including steroids, hemoglobin doping, neurostimulation, and CRISPR. Steroids have become a widespread method for enhancing muscle mass in strength training. It activates signaling proteins like mTOR, Akt, etc., and alters protein synthesis pathways, cell cycle, oxidative stress, and apoptosis, promoting muscle growth(Wenbo and Yan). Likewise, hemoglobin doping involves either direct blood transfusion or indirect methods of increasing hemoglobin via stimulating erythropoiesis. These mechanisms increase the oxygen-delivering capacity to active tissues(Handelsman). Additionally, neurostimulation tools such as transcranial direct current stimulation, transcranial random noise stimulation, transcranial alternating current stimulation, transcranial magnetic stimulation, and transcutaneous vagus nerve stimulation could potentially modify cerebral, mental, and physical performance(Brain Sciences | Special Issue : Neurostimulation Techniques and Physical Exercise). By applying electrical currents to achieve functional activation or inhibition of specific neuronal pathways, neurostimulation could enhance cognitive communications(Neurostimulation - an Overview | ScienceDirect Topics).

However, these biological enhancements for athletic performance have multiple limitations. While anabolic steroids impact sports performance, it also poses significant health risks and psychological consequences. This is not limited to increased risk of cardiovascular-related diseases, infertility, liver dysfunction, aggression, and anxiety(Bhasin et al.). Hemoglobin doping may raise hematocrit, which may increase the viscosity and hypercoagulability of blood, running a high risk of venous thrombosis and pulmonary embolism if the athlete spends many hours relatively immobile(Jones and Tunstall Pedoe). Moreover, a large disconnect remains between how each neuromodulation approach affects single cells and how these perturbations translate to network function and ultimately to behavior(Johnson et al.). Thus, it is difficult to identify and target the mechanism of choice for enhancing performance.

With the advent of gene editing, attention is now drawn to methods to correct genetic disorders that were previously not clinically tractable, such as Duchenne muscular dystrophy (DMD). DMD is a hereditary degenerative muscle condition that severely impacts the motor

ability of children, causing weakness in the skeletal, diaphragm, and cardiac muscles(Heydemann and Siemionow). Recently, a gene-therapy-based approach was approved in the US to treat DMD, Dystrophin gene (Hoy). The therapeutic technology leverages adeno-associated virus (AAV) technology to deliver a gene encoding a micro-dystrophin protein. This is a shortened (138 kDa) version of the dystrophin protein expressed in normal muscle cells(Hoy), which restores genetic function in muscle cells.

Similarly in 2023, the first CRISPR-based gene therapy was approved in the US and Europe for treating patients with transfusion-dependent β -thalassemia and treating sickle cell disease in patients aged ≥ 12 years with recurrent vaso-occlusive crises(Parums). As such, these breakthroughs indicate key organ systems, muscles, and blood, can be effectively targeted and modulated through gene therapy and gene editing. As such there could be potential methods for modulating athletic performance. For example, CRISPR is a highly effective gene editing tool that uses the defense system of certain bacteria against viruses and plasmids for therapeutic purposes(Rodríguez-Rodríguez et al.). CRISPR-Cas systems use a guide RNA molecule to locate and act on a specific genetic locus(Tournas et al.). It reduces the risk of off-target edits compared to other gene editing methods. However, it is currently against the guidelines of the World Anti-Doping Agency(2024list_en_final_22_september_2023.Pdf).

As such, employing such methods would be unethical at present. For example, doping in sports is considered cheating and unfair because it harms the athletes, and it is unnatural and dehumanizing(Schneider and Friedmann). It jeopardizes an athlete's health and undermines competitive fairness(Palmi et al.).

In conclusion, while this paper provides an overview of genes that influence athletic performance, additional research is required to deepen our understanding of how athletes can enhance their strength, power, and mental capabilities and the key ethical decisions surrounding genetics research in this space.

Works Cited

2024list_en_final_22_september_2023.Pdf.

https://www.wada-ama.org/sites/default/files/2023-09/2024list_en_final_22_september_2023.pdf.
Accessed 25 June 2024.

ACE Angiotensin I Converting Enzyme [Homo Sapiens (Human)] - Gene - NCBI.

<https://www.ncbi.nlm.nih.gov/gene/1636>. Accessed 19 May 2024.

ACTN3 - an Overview | ScienceDirect Topics.

<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/actn3>. Accessed 20 May 2024.

ACTN3 Actinin Alpha 3 [Homo Sapiens (Human)] - Gene - NCBI.

<https://www.ncbi.nlm.nih.gov/gene/89>. Accessed 19 May 2024.

ADRB1 Adrenoceptor Beta 1 [Homo Sapiens (Human)] - Gene - NCBI.

<https://www.ncbi.nlm.nih.gov/gene/153>. Accessed 26 May 2024.

ADRB2 Adrenoceptor Beta 2 [Homo Sapiens (Human)] - Gene - NCBI.

<https://www.ncbi.nlm.nih.gov/gene/154>. Accessed 19 May 2024.

Aglioti, Salvatore, et al. "Action Anticipation and Motor Resonance in Elite Basketball Players."

Nature Neuroscience, vol. 11, Oct. 2008, pp. 1109–16. ResearchGate, <https://doi.org/10.1038/nn.2182>.

Ahlgrim, Christoph, et al. "Applying the Optimized CO Rebreathing Method for Measuring

Blood Volumes and Hemoglobin Mass in Heart Failure Patients." *Frontiers in Physiology*,
vol. 9, Nov. 2018, p. 1603. PubMed Central, <https://doi.org/10.3389/fphys.2018.01603>.

Ahmad Yusof, Hazwani, and Ahmad Munir Che Muhamed. "Angiotensin-Converting Enzyme
(ACE) Insertion/Deletion Gene Polymorphism across Ethnicity: A Narrative Review of
Performance Gene." *Sport Sciences for Health*, vol. 17, no. 1, Mar. 2021, pp. 57–77.

Springer Link, <https://doi.org/10.1007/s11332-020-00712-9>.

Ahmetov, I. I., A. M. Druzhevskaya, et al. "The ACTN3 R577X Polymorphism in Russian
Endurance Athletes." *British Journal of Sports Medicine*, vol. 44, no. 9, July 2010, pp.
649–52. *bjsm.bmj.com*, <https://doi.org/10.1136/bjsm.2008.051540>.

Ahmetov, I. I., D. V. Popov, et al. "The Use of Molecular Genetic Methods for Prognosis of
Aerobic and Anaerobic Performance in Athletes." *Human Physiology*, vol. 34, no. 3, May
2008, pp. 338–42. Springer Link, <https://doi.org/10.1134/S0362119708030110>.

Ahmetov, Ildus I, et al. "PPAR α Gene Variation and Physical Performance in Russian Athletes."
European Journal of Applied Physiology, vol. 97, no. 1, May 2006, pp. 103–08. Springer
Link, <https://doi.org/10.1007/s00421-006-0154-4>.

Ahmetov, Ildus I., et al. "The Combined Impact of Metabolic Gene Polymorphisms on Elite
Endurance Athlete Status and Related Phenotypes." *Human Genetics*, vol. 126, no. 6,
Dec. 2009, pp. 751–61. Springer Link, <https://doi.org/10.1007/s00439-009-0728-4>.

Akhmetov: Association of Regulatory Genes Polymorphisms... - Google Scholar.

https://scholar.google.com/scholar_lookup?title=Association+of+regulatory+genes+polymorphisms+with+aerobic+and+anaerobic+performance+of+athletes&author=Akhmetov,+I.I.&author=Popov,+D.V.&author=Mozha%C4%ADskaia,+I.A.&author=Missina,+S.S.&author=Astratenkova,+I.V.&author=Vinogradova,+O.L.&author=Rogozkin,+V.A.&publication_year=2007&journal=Ross.+Fiziol.+Zh.+Im.+I.M.+Sechenova&volume=93&pages=837%E2%80%93843. Accessed 19 May 2024.

Alvarez, Ruth, et al. "Genetic Variation in the Renin-Angiotensin System and Athletic
Performance." *European Journal of Applied Physiology*, vol. 82, no. 1, May 2000, pp.

117–20. Springer Link, <https://doi.org/10.1007/s004210050660>.

AMPD1 Adenosine Monophosphate Deaminase 1 [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/270>. Accessed 19 May 2024.

Angiotensin-Converting Enzyme 2 - an Overview | ScienceDirect Topics. <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/angiotensin-converting-enzyme-2>. Accessed 20 May 2024.

Ash, Garrett I., et al. “No Association between ACE Gene Variation and Endurance Athlete Status in Ethiopians.” *Medicine & Science in Sports & Exercise*, vol. 43, no. 4, Apr. 2011, p. 590. journals.lww.com, <https://doi.org/10.1249/MSS.0b013e3181f70bd6>.

Barbalato, Laura, and Leela Sharath Pillarisetty. “Histology, Red Blood Cell.” StatPearls, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK539702/>.

Bathina, Siresha, and Undurti N. Das. “Brain-Derived Neurotrophic Factor and Its Clinical Implications.” *Archives of Medical Science : AMS*, vol. 11, no. 6, Dec. 2015, pp. 1164–78. PubMed Central, <https://doi.org/10.5114/aoms.2015.56342>.

BDKRB2 Bradykinin Receptor B2 [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/624>. Accessed 19 May 2024.

Bhasin, Shalender, et al. “Anabolic-Androgenic Steroid Use in Sports, Health, and Society.” *Medicine & Science in Sports & Exercise*, vol. 53, no. 8, Aug. 2021, p. 1778. journals.lww.com, <https://doi.org/10.1249/MSS.0000000000002670>.

Brain Sciences | Special Issue : Neurostimulation Techniques and Physical Exercise. https://www.mdpi.com/journal/brainsci/special_issues/NE2S673Q1K. Accessed 23 June 2024.

Byrnes, James R., and Alisa S. Wolberg. “Red Blood Cells in Thrombosis.” *Blood*, vol. 130, no. 16, Oct. 2017, pp. 1795–99. PubMed Central, <https://doi.org/10.1182/blood-2017-03-745349>.

Chakrapani, Sumita, et al. “Neuroplasticity and the Biological Role of Brain Derived Neurotrophic Factor in the Pathophysiology and Management of Depression.” *Cureus*, vol. 12, no. 11, p. e11396. PubMed Central, <https://doi.org/10.7759/cureus.11396>.

Chicharro, J. L., et al. “Mutations in the Hereditary Haemochromatosis Gene HFE in Professional Endurance Athletes.” *British Journal of Sports Medicine*, vol. 38, no. 4, Aug. 2004, pp. 418–21. bjsm.bmj.com, <https://doi.org/10.1136/bjsm.2002.003921>.

Cięszczyk, P., et al. “Is the C34T Polymorphism of the AMPD1 Gene Associated with Athlete Performance in Rowing?” *International Journal of Sports Medicine*, vol. 32, no. 12, Dec. 2011, pp. 987–91. www.thieme-connect.de, <https://doi.org/10.1055/s-0031-1283186>.

Cieszczyk, P., et al. “The Angiotensin Converting Enzyme Gene I/D Polymorphism in Polish Rowers.” *International Journal of Sports Medicine*, vol. 30, no. 08, Aug. 2009, pp. 624–27. www.thieme-connect.de, <https://doi.org/10.1055/s-0029-1202825>.

Collins, Malcolm, et al. “The ACE Gene and Endurance Performance during the South African Ironman Triathlons.” *Medicine & Science in Sports & Exercise*, vol. 36, no. 8, Aug. 2004, p. 1314. journals.lww.com, <https://doi.org/10.1249/01.MSS.0000135779.41475.42>.

COMT Catechol-O-Methyltransferase [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/1312>. Accessed 26 May 2024.

Davidovits, Paul. “Chapter 13 - Electricity.” *Physics in Biology and Medicine (Fifth Edition)*,

- edited by Paul Davidovits, Academic Press, 2019, pp. 193–211. ScienceDirect, <https://doi.org/10.1016/B978-0-12-813716-1.00013-6>.
- De Moor, Marleen H. M., et al. “Genome-Wide Linkage Scan for Athlete Status in 700 British Female DZ Twin Pairs.” *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, vol. 10, no. 6, Dec. 2007, pp. 812–20. PubMed, <https://doi.org/10.1375/twin.10.6.812>.
- D’Errico, Ilenia, et al. “Peroxisome Proliferator-Activated Receptor- γ Coactivator 1- α (PGC1 α) Is a Metabolic Regulator of Intestinal Epithelial Cell Fate.” *Proceedings of the National Academy of Sciences*, vol. 108, no. 16, Apr. 2011, pp. 6603–08. pnas.org (Atypon), <https://doi.org/10.1073/pnas.1016354108>.
- Desvergne, Béatrice, and Walter Wahli. “Peroxisome Proliferator-Activated Receptors: Nuclear Control of Metabolism*.” *Endocrine Reviews*, vol. 20, no. 5, Oct. 1999, pp. 649–88. Silverchair, <https://doi.org/10.1210/edrv.20.5.0380>.
- Deugnier, Yves, et al. “Increased Body Iron Stores in Elite Road Cyclists.” *Medicine & Science in Sports & Exercise*, vol. 34, no. 5, May 2002, p. 876.
- Döring, Frank, et al. “A Common Haplotype and the Pro582Ser Polymorphism of the Hypoxia-Inducible Factor-1 α (HIF1A) Gene in Elite Endurance Athletes.” *Journal of Applied Physiology*, vol. 108, no. 6, June 2010, pp. 1497–500. journals.physiology.org (Atypon), <https://doi.org/10.1152/jappphysiol.01165.2009>.
- Döring, Frank E., et al. “ACTN3 R577X and Other Polymorphisms Are Not Associated with Elite Endurance Athlete Status in the Genathlete Study.” *Journal of Sports Sciences*, vol. 28, no. 12, Oct. 2010, pp. 1355–59. tandfonline.com (Atypon), <https://doi.org/10.1080/02640414.2010.507675>.
- Douglas, P. S., and M. O’Toole. “Aging and Physical Activity Determine Cardiac Structure and Function in the Older Athlete.” *Journal of Applied Physiology* (Bethesda, Md.: 1985), vol. 72, no. 5, May 1992, pp. 1969–73. PubMed, <https://doi.org/10.1152/jappl.1992.72.5.1969>.
- DRD2 Dopamine Receptor D2 [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/1813>. Accessed 26 May 2024.
- Drozdovska: Allelic Polymorphism of Endothelial NO-Syntha... - Google Scholar. [https://scholar.google.com/scholar_lookup?title=Allelic+Polymorphism+of+Endothelial+No-Synthase+\(eNOS\)+Association+with+Exercise-Induced+Hypoxia+Adaptation&author=Drozdovska,+S.B.&author=Dosenko,+V.E.&author=Ilyin,+V.N.&author=Filipov,+M.M.&author=Kuzmina,+L.M.&publication_year=2009&journal=Baltic.+J.+Health+Phys.+Activ.&volume=1&pages=13%E2%80%9319&doi=10.2478/v10131-009-0001-1](https://scholar.google.com/scholar_lookup?title=Allelic+Polymorphism+of+Endothelial+No-Synthase+(eNOS)+Association+with+Exercise-Induced+Hypoxia+Adaptation&author=Drozdovska,+S.B.&author=Dosenko,+V.E.&author=Ilyin,+V.N.&author=Filipov,+M.M.&author=Kuzmina,+L.M.&publication_year=2009&journal=Baltic.+J.+Health+Phys.+Activ.&volume=1&pages=13%E2%80%9319&doi=10.2478/v10131-009-0001-1). Accessed 19 May 2024.
- Eider, J., et al. “The Association between D Allele of the ACE Gene and Power Performance in Polish Elite Athletes.” *Science & Sports*, vol. 28, no. 6, Dec. 2013, pp. 325–30. ScienceDirect, <https://doi.org/10.1016/j.scispo.2012.11.005>.
- FEV FEV Transcription Factor, ETS Family Member [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/54738>. Accessed 26 May 2024.
- Fountain, John H., et al. “Physiology, Renin Angiotensin System.” StatPearls, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK470410/>.
- Gangat, Naseema, et al. “JAK2 Unmutated Erythrocytosis: 2023 Update on Diagnosis and Management.” *American Journal of Hematology*, vol. 98, no. 6, 2023, pp. 965–81. Wiley Online Library, <https://doi.org/10.1002/ajh.26920>.

- Gasser, Benedikt, et al. “Association of Gene Variants for Mechanical and Metabolic Muscle Quality with Cardiorespiratory and Muscular Variables Related to Performance in Skiing Athletes.” *Genes*, vol. 13, no. 10, 10, Oct. 2022, p. 1798. www.mdpi.com, <https://doi.org/10.3390/genes13101798>.
- Gavin, Rory. *Factors That Effect Sports Performance: Sleep, Mood, Having a Pre-Match Routine*.
- Gayagay, George, et al. “Elite Endurance Athletes and the ACE I Allele – the Role of Genes in Athletic Performance.” *Human Genetics*, vol. 103, no. 1, July 1998, pp. 48–50. Springer Link, <https://doi.org/10.1007/s004390050781>.
- Geisler, Maria, et al. “Brains of Endurance Athletes Differ in the Association Areas but Not in the Primary Areas.” *Psychophysiology*, vol. 61, no. 4, 2024, p. e14483. Wiley Online Library, <https://doi.org/10.1111/psyp.14483>.
- Ginevičienė: Relating Fitness Phenotypes to Genotypes... - Google Scholar.
https://scholar.google.com/scholar_lookup?title=Relating+fitness+phenotypes+to+genotypes+in+Lithuanian+elite+athletes&author=Ginevi%C4%8Dien%C4%97,+V.&author=Pranckevi%C4%8Dien%C4%97,+E.&author=Mila%C5%A1ius,+K.&author=Ku%C4%8Dinskas,+V.&publication_year=2010&journal=Acta+Med.+Litu.&volume=17&pages=1%E2%80%9310&doi=10.2478/v10140-010-0001-0. Accessed 19 May 2024.
- Ginevičienė, Valentina, Audronė Jakaitienė, et al. “AMPD1rs17602729 Is Associated with Physical Performance of Sprint and Power in Elite Lithuanian Athletes.” *BMC Genetics*, vol. 15, no. 1, May 2014, p. 58. BioMed Central, <https://doi.org/10.1186/1471-2156-15-58>.
- Ginevičienė, Valentina, Aidas Pranculis, et al. “Genetic Variation of the Human ACE and ACTN3 Genes and Their Association With Functional Muscle Properties in Lithuanian Elite Athletes.” *Medicina*, vol. 47, no. 5, 5, May 2011, p. 40. www.mdpi.com, <https://doi.org/10.3390/medicina47050040>.
- Gómez-Gallego, Félix, et al. “The –786 T/C Polymorphism of the NOS3 Gene Is Associated with Elite Performance in Power Sports.” *European Journal of Applied Physiology*, vol. 107, no. 5, Nov. 2009, pp. 565–69. Springer Link, <https://doi.org/10.1007/s00421-009-1166-7>.
- Grealy, Rebecca, et al. “The Genetics of Endurance: Frequency of the ACTN3 R577X Variant in Ironman World Championship Athletes.” *Journal of Science and Medicine in Sport*, vol. 16, no. 4, July 2013, pp. 365–71. www.jsams.org, <https://doi.org/10.1016/j.jsams.2012.08.013>.
- Grenda: Bdkrb2 Gene-9/+ 9 Polymorphism and Swimming... - Google Scholar.
https://scholar.google.com/scholar_lookup?title=Bdkrb2+gene+-9/+9+polymorphism+and+swimming+performance&author=Grenda,+A.&author=Leo%C5%84ska-Duniec,+A.&author=Ci%C4%99szyk,+P.&author=Zmijewski,+P.&publication_year=2014&journal=Biol.+Sport&volume=31&pages=109%E2%80%93113&doi=10.5604/20831862.1096047. Accessed 19 May 2024.
- Guilherme, João Paulo L. F., et al. “The BDNF-Increasing Allele Is Associated With Increased Proportion of Fast-Twitch Muscle Fibers, Handgrip Strength, and Power Athlete Status.” *Journal of Strength and Conditioning Research*, vol. 36, no. 7, July 2022, pp. 1884–89. DOI.org (Crossref), <https://doi.org/10.1519/JSC.0000000000003756>.
- Hafen, Brant B., et al. “Anatomy, Smooth Muscle.” *StatPearls*, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK532857/>.
- Hall, Elliott C. R., et al. “The PPARGC1A Gly482Ser Polymorphism Is Associated with Elite Long-Distance Running Performance.” *Journal of Sports Sciences*, vol. 41, no. 1, Jan.

- 2023, pp. 56–62. Taylor and Francis+NEJM, <https://doi.org/10.1080/02640414.2023.2195737>.
- Handelsman, David J. “Performance Enhancing Hormone Doping in Sport.” Endotext, edited by Kenneth R. Feingold et al., MDText.com, Inc., 2000. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK305894/>.
- Harber, M. P., et al. “Single Muscle Fiber Contractile Properties during a Competitive Season in Male Runners.” *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, vol. 287, no. 5, Nov. 2004, pp. R1124-1131. PubMed, <https://doi.org/10.1152/ajpregu.00686.2003>.
- Harber, Matthew, and Scott Trappe. “Single Muscle Fiber Contractile Properties of Young Competitive Distance Runners.” *Journal of Applied Physiology (Bethesda, Md.: 1985)*, vol. 105, no. 2, Aug. 2008, pp. 629–36. PubMed, <https://doi.org/10.1152/jappphysiol.00995.2007>.
- He, Z. H., et al. “PGC-Related Gene Variants and Elite Endurance Athletic Status in a Chinese Cohort: A Functional Study.” *Scandinavian Journal of Medicine & Science in Sports*, vol. 25, no. 2, 2015, pp. 184–95. Wiley Online Library, <https://doi.org/10.1111/sms.12188>.
- Heinicke, K., et al. “Blood Volume and Hemoglobin Mass in Elite Athletes of Different Disciplines.” *International Journal of Sports Medicine*, vol. 22, no. 7, Oct. 2001, pp. 504–12. PubMed, <https://doi.org/10.1055/s-2001-17613>.
- Hermine, Olivier, et al. “Eighty Percent of French Sport Winners in Olympic, World and Europeans Competitions Have Mutations in the Hemochromatosis HFE Gene.” *Biochimie*, vol. 119, Dec. 2015, pp. 1–5. ScienceDirect, <https://doi.org/10.1016/j.biochi.2015.09.028>.
- Heydemann, Ahlke, and Maria Siemionow. “A Brief Review of Duchenne Muscular Dystrophy Treatment Options, with an Emphasis on Two Novel Strategies.” *Biomedicines*, vol. 11, no. 3, Mar. 2023, p. 830. PubMed Central, <https://doi.org/10.3390/biomedicines11030830>.
- HFE Homeostatic Iron Regulator [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/3077>. Accessed 19 May 2024.
- HIF1A Hypoxia Inducible Factor 1 Subunit Alpha [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/3091>. Accessed 19 May 2024.
- Hoy, Sheridan M. “Delandistrogene Moxeparvec: First Approval.” *Drugs*, vol. 83, no. 14, Sept. 2023, pp. 1323–29. PubMed, <https://doi.org/10.1007/s40265-023-01929-x>.
- Increased Cortical Thickness in Sports Experts: A Comparison of Diving Players with the Controls | PLOS ONE. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017112>. Accessed 25 May 2024.
- Introduction to the Muscular System | SEER Training. <https://training.seer.cancer.gov/anatomy/muscular/>. Accessed 10 May 2024.
- Jacob, Ysabel, et al. “Genetic Predictors of Match Performance in Sub-Elite Australian Football Players: A Pilot Study.” *Journal of Exercise Science & Fitness*, vol. 17, no. 2, May 2019, pp. 41–46. ScienceDirect, <https://doi.org/10.1016/j.jesf.2018.10.007>.
- Jäncke, Lutz, et al. “The Architecture of the Golfer’s Brain.” *PLoS ONE*, vol. 4, no. 3, Mar. 2009, p. e4785. PubMed Central, <https://doi.org/10.1371/journal.pone.0004785>.
- Jelaković, B., et al. “I022: Influence of Angiotensin Converting Enzyme (ACE) Gene Polymorphism and Circadian Blood Pressure (BP) Changes on Left Ventricle (LV) Mass

- in Competitive Oarsmen.” *American Journal of Hypertension*, vol. 13, no. S2, Apr. 2000, p. 182A. Silverchair, [https://doi.org/10.1016/S0895-7061\(00\)01168-7](https://doi.org/10.1016/S0895-7061(00)01168-7).
- Ji, Zinan. “The Ultimate Athlete: Genetics Vs. Training.” *SHS Web of Conferences*, edited by J. Rak, vol. 157, 2023, p. 04017. DOI.org (Crossref), <https://doi.org/10.1051/shsconf/202315704017>.
- Johnson, Matthew D., et al. “Neuromodulation for Brain Disorders: Challenges and Opportunities.” *IEEE Transactions on Bio-Medical Engineering*, vol. 60, no. 3, Mar. 2013, pp. 610–24. PubMed Central, <https://doi.org/10.1109/TBME.2013.2244890>.
- Jones, M., and D. S. Tunstall Pedoe. “Blood Doping--a Literature Review.” *British Journal of Sports Medicine*, vol. 23, no. 2, June 1989, pp. 84–88.
- Juvonen, E., et al. “Autosomal Dominant Erythrocytosis Caused by Increased Sensitivity to Erythropoietin.” *Blood*, vol. 78, no. 11, Dec. 1991, pp. 3066–69.
- Keohane, C., et al. “The Diagnosis and Management of Erythrocytosis.” *BMJ*, vol. 347, no. nov18 1, Nov. 2013, pp. f6667–f6667. DOI.org (Crossref), <https://doi.org/10.1136/bmj.f6667>.
- Kitazawa, Hiroya, et al. “Potential Genetic Contributions of the Central Nervous System to a Predisposition to Elite Athletic Traits: State-of-the-Art and Future Perspectives.” *Genes*, vol. 12, no. 3, Mar. 2021. www.ncbi.nlm.nih.gov, <https://doi.org/10.3390/genes12030371>.
- Kumagai, Hiroshi, et al. “Role of Selected Polymorphisms in Determining Muscle Fiber Composition in Japanese Men and Women.” *Journal of Applied Physiology (Bethesda, Md.: 1985)*, vol. 124, no. 5, May 2018, pp. 1377–84. PubMed, <https://doi.org/10.1152/jappphysiol.00953.2017>.
- Lack of Association between the GNB3 Rs5443, HIF1A Rs11549465 Polymorphisms, Physiological and Functional Characteristics - Bosnyák - 2020 - *Annals of Human Genetics* - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1111/ahg.12387>. Accessed 19 May 2024.
- Lucia, A., et al. “ACTN3 Genotype in Professional Endurance Cyclists.” *International Journal of Sports Medicine*, vol. 27, no. 11, Nov. 2006, pp. 880–84. www.thieme-connect.de, <https://doi.org/10.1055/s-2006-923862>.
- Lucía, A., et al. “Is There an Association between ACE and CKMM Polymorphisms and Cycling Performance Status during 3-Week Races?” *International Journal of Sports Medicine*, vol. 26, no. 06, July 2005, pp. 442–47. www.thieme-connect.de, <https://doi.org/10.1055/s-2004-821108>.
- Lucia: PPARGC1A Genotype (Gly482Ser) Predicts Exceptional... - Google Scholar. [https://scholar.google.com/scholar_lookup?title=PPARGC1A+genotype+\(Gly482Ser\)+predicts+exceptional+endurance+capacity+in+European+men&author=Lucia,+A.&author=G%C3%B3mez-Gallego,+F.&author=Barroso,+I.&author=Rabad%C3%A1n,+M.&author=Bandr%C3%A9s,+F.&author=San+Juan,+A.F.&author=Chicharro,+J.L.&author=Ekelund,+U.&author=Brage,+S.&author=Earnest,+C.P.&publication_year=2005&journal=J.+Appl.+Physiol.&volume=99&pages=344%E2%80%93348&doi=10.1152/jappphysiol.00037.2005](https://scholar.google.com/scholar_lookup?title=PPARGC1A+genotype+(Gly482Ser)+predicts+exceptional+endurance+capacity+in+European+men&author=Lucia,+A.&author=G%C3%B3mez-Gallego,+F.&author=Barroso,+I.&author=Rabad%C3%A1n,+M.&author=Bandr%C3%A9s,+F.&author=San+Juan,+A.F.&author=Chicharro,+J.L.&author=Ekelund,+U.&author=Brage,+S.&author=Earnest,+C.P.&publication_year=2005&journal=J.+Appl.+Physiol.&volume=99&pages=344%E2%80%93348&doi=10.1152/jappphysiol.00037.2005). Accessed 19 May 2024.
- Ma, Chongyang, et al. “Erythropoietin Pathway: A Potential Target for the Treatment of Depression.” *International Journal of Molecular Sciences*, vol. 17, no. 5, May 2016, p. 677. PubMed Central, <https://doi.org/10.3390/ijms17050677>.
- Ma, Fang, et al. “The Association of Sport Performance with ACE and ACTN3 Genetic Polymorphisms: A Systematic Review and Meta-Analysis.” *PLoS ONE*, vol. 8, no. 1, Jan. 2013, p. e54685. PubMed Central, <https://doi.org/10.1371/journal.pone.0054685>.
- Maciejewska, Agnieszka, Marek Sawczuk, Pawel Cieszczyk, et al. “The PPARGC1A gene

- Gly482Ser in Polish and Russian Athletes.” *Journal of Sports Sciences*, vol. 30, no. 1, Jan. 2012, pp. 101–13. Taylor and Francis+NEJM, <https://doi.org/10.1080/02640414.2011.623709>.
- Maciejewska, Agnieszka, Marek Sawczuk, and Paweł Cieższyk. “Variation in the PPAR α Gene in Polish Rowers.” *Journal of Science and Medicine in Sport*, vol. 14, no. 1, Jan. 2011, pp. 58–64. www.jsams.org, <https://doi.org/10.1016/j.jsams.2010.05.006>.
- Maharam, L., et al. “Masters Athletes: Factors Affecting Performance.” *Sports Medicine (Auckland, N.Z.)*, vol. 28, Nov. 1999, pp. 273–85.
- Malczewska-Lenczowska, Jadwiga, et al. “Total Hemoglobin Mass, Aerobic Capacity, and HBB Gene in Polish Road Cyclists.” *The Journal of Strength & Conditioning Research*, vol. 30, no. 12, Dec. 2016, p. 3512. journals.lww.com, <https://doi.org/10.1519/JSC.0000000000001435>.
- McCuller, Christopher, et al. “Physiology, Skeletal Muscle.” *StatPearls*, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK537139/>.
- Mikami, E., et al. “ACTN3 R577X Genotype Is Associated with Sprinting in Elite Japanese Athletes.” *International Journal of Sports Medicine*, vol. 35, no. 02, Feb. 2014, pp. 172–77. www.thieme-connect.de, <https://doi.org/10.1055/s-0033-1347171>.
- Min, Seok-Ki, et al. “Is There a Gender Difference between ACE Gene and Race Distance?” *Applied Physiology, Nutrition, and Metabolism*, vol. 34, no. 5, Oct. 2009, pp. 926–32. cdnsiencepub.com (Atypon), <https://doi.org/10.1139/H09-097>.
- Miranda, Magdalena, et al. “Brain-Derived Neurotrophic Factor: A Key Molecule for Memory in the Healthy and the Pathological Brain.” *Frontiers in Cellular Neuroscience*, vol. 13, Aug. 2019. *Frontiers*, <https://doi.org/10.3389/fncel.2019.00363>.
- Mitochondrial DNA Variation Is Associated with Elite Athletic Status in the Polish Population - Maruszak - 2014 - *Scandinavian Journal of Medicine & Science in Sports* - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1111/sms.12012>. Accessed 19 May 2024.
- Montgomery, H. E., et al. “Human Gene for Physical Performance.” *Nature*, vol. 393, no. 6682, May 1998, pp. 221–22. www.nature.com, <https://doi.org/10.1038/30374>.
- Müller, Sean, et al. “How Do World-Class Cricket Batsmen Anticipate a Bowler’s Intention?” *The Quarterly Journal of Experimental Psychology*, Dec. 2006. *world*, www.tandfonline.com, <https://doi.org/10.1080/02643290600576595>.
- Muniesa, Carlos A., et al. “World-Class Performance in Lightweight Rowing: Is It Genetically Influenced? A Comparison with Cyclists, Runners and Non-Athletes.” *British Journal of Sports Medicine*, vol. 44, no. 12, Sept. 2010, pp. 898–901. bjsm.bmj.com, <https://doi.org/10.1136/bjsm.2008.051680>.
- Muñoz Ospina, Beatriz, and Natalia Cadavid-Ruiz. “The Effect of Aerobic Exercise on Serum Brain-Derived Neurotrophic Factor (BDNF) and Executive Function in College Students.” *Mental Health and Physical Activity*, vol. 26, Mar. 2024, p. 100578. *ScienceDirect*, <https://doi.org/10.1016/j.mhpa.2024.100578>.
- Muscle Types | SEER Training. <https://training.seer.cancer.gov/anatomy/muscular/types.html>. Accessed 10 May 2024.
- Mustafina, Leysan J., et al. “AGTR2 Gene Polymorphism Is Associated with Muscle Fibre

- Composition, Athletic Status and Aerobic Performance.” *Experimental Physiology*, vol. 99, no. 8, Aug. 2014, pp. 1042–52. PubMed, <https://doi.org/10.1113/expphysiol.2014.079335>.
- Myerson, Saul, et al. “Human Angiotensin I-Converting Enzyme Gene and Endurance Performance.” *Journal of Applied Physiology*, vol. 87, no. 4, Oct. 1999, pp. 1313–16. Semantic Scholar, <https://doi.org/10.1152/jappl.1999.87.4.1313>.
- Nazarov, Igor B., et al. “The Angiotensin Converting Enzyme I/D Polymorphism in Russian Athletes.” *European Journal of Human Genetics*, vol. 9, no. 10, Oct. 2001, pp. 797–801. www.nature.com, <https://doi.org/10.1038/sj.ejhg.5200711>.
- Neurostimulation - an Overview | ScienceDirect Topics.
<https://www.sciencedirect.com/topics/medicine-and-dentistry/neurostimulation>. Accessed 23 June 2024.
- Niemi, Anna-Kaisa, and Kari Majamaa. “Mitochondrial DNA and ACTN3 Genotypes in Finnish Elite Endurance and Sprint Athletes.” *European Journal of Human Genetics*, vol. 13, no. 8, Aug. 2005, pp. 965–69. www.nature.com, <https://doi.org/10.1038/sj.ejhg.5201438>.
- Noohi, Fatemeh, et al. “Association of COMT Val158met and DRD2 G>T Genetic Polymorphisms with Individual Differences in Motor Learning and Performance in Female Young Adults.” *Journal of Neurophysiology*, vol. 111, no. 3, Feb. 2014, pp. 628–40. PubMed, <https://doi.org/10.1152/jn.00457.2013>.
- Normal Red Blood Cells’ Shape Stabilized by Membrane’s in-Plane Ordering | Scientific Reports. <https://www.nature.com/articles/s41598-019-56128-0>. Accessed 8 June 2024.
- NOS3 Nitric Oxide Synthase 3 [Homo Sapiens (Human)] - Gene - NCBI.
<https://www.ncbi.nlm.nih.gov/gene/4846>. Accessed 19 May 2024.
- Orysiak, Joanna, et al. “Overrepresentation of the ACTN3 XX Genotype in Elite Canoe and Kayak Paddlers.” *The Journal of Strength & Conditioning Research*, vol. 29, no. 4, Apr. 2015, p. 1107. journals.lww.com, <https://doi.org/10.1519/JSC.0000000000000717>.
- Overney, Leila S., et al. “Enhanced Temporal but Not Attentional Processing in Expert Tennis Players.” *PLoS ONE*, vol. 3, no. 6, June 2008, p. e2380. PubMed Central, <https://doi.org/10.1371/journal.pone.0002380>.
- Palmi, I., et al. “The Unethicality of Doping in Sports.” *La Clinica Terapeutica*, vol. 170, no. 2, 2019, pp. e100–01. PubMed, <https://doi.org/10.7417/CT.2019.2117>.
- Papadimitriou, I. D., et al. “The ACTN3 Gene in Elite Greek Track and Field Athletes.” *International Journal of Sports Medicine*, vol. 29, no. 04, Apr. 2008, pp. 352–55. www.thieme-connect.de, <https://doi.org/10.1055/s-2007-965339>.
- Papadimitriou, Ioannis D., Alejandro Lucia, et al. “ACTN3 R577X and ACE I/D Gene Variants Influence Performance in Elite Sprinters: A Multi-Cohort Study.” *BMC Genomics*, vol. 17, no. 1, Apr. 2016, p. 285. BioMed Central, <https://doi.org/10.1186/s12864-016-2462-3>.
- Papadimitriou, Ioannis D., Sarah J. Lockey, et al. “No Association between ACTN3 R577X and ACE I/D Polymorphisms and Endurance Running Times in 698 Caucasian Athletes.” *BMC Genomics*, vol. 19, no. 1, Jan. 2018, p. 13. BioMed Central, <https://doi.org/10.1186/s12864-017-4412-0>.
- Paparini, Andrea, et al. “ACTN3 Genotyping by Real-Time PCR in the Italian Population and Athletes.” *Medicine & Science in Sports & Exercise*, vol. 39, no. 5, May 2007, p. 810.

- journals.lww.com, <https://doi.org/10.1097/mss.0b013e3180317491>.
- Park, In Sung, et al. “White Matter Plasticity in the Cerebellum of Elite Basketball Athletes.” *Anatomy & Cell Biology*, vol. 48, no. 4, Dec. 2015, pp. 262–67. PubMed Central, <https://doi.org/10.5115/acb.2015.48.4.262>.
- Paruk, T., et al. “Structural Brain Differences between Ultra-Endurance Athletes and Sedentary Persons.” *Sports Medicine and Health Science*, vol. 2, no. 2, June 2020, pp. 89–94. ScienceDirect, <https://doi.org/10.1016/j.smhs.2020.05.004>.
- Parums, Dinah V. “Editorial: First Regulatory Approvals for CRISPR-Cas9 Therapeutic Gene Editing for Sickle Cell Disease and Transfusion-Dependent β -Thalassemia.” *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, vol. 30, Mar. 2024, pp. e944204-1-e944204-4. PubMed Central, <https://doi.org/10.12659/MSM.944204>.
- Peplonska, Beata, et al. “Association of Serotonergic Pathway Gene Variants with Elite Athletic Status in the Polish Population.” *Journal of Sports Sciences*, vol. 37, no. 14, July 2019, pp. 1655–62. Taylor and Francis+NEJM, <https://doi.org/10.1080/02640414.2019.1583156>.
- Pittman, Roland N. “The Circulatory System and Oxygen Transport.” *Regulation of Tissue Oxygenation*, Morgan & Claypool Life Sciences, 2011. www.ncbi.nlm.nih.gov, <https://www.ncbi.nlm.nih.gov/books/NBK54112/>.
- PPARA Peroxisome Proliferator Activated Receptor Alpha [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/5465>. Accessed 19 May 2024.
- PPARGC1A PPARG Coactivator 1 Alpha [Homo Sapiens (Human)] - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/10891>. Accessed 19 May 2024.
- Prior, Steven J., et al. “Sequence Variation in Hypoxia-Inducible Factor 1 α (HIF1A): Association with Maximal Oxygen Consumption.” *Physiological Genomics*, vol. 15, no. 1, Sept. 2003, pp. 20–26. Semantic Scholar, <https://doi.org/10.1152/physiolgenomics.00061.2003>.
- PubChem. BDNF - Brain Derived Neurotrophic Factor (Human). <https://pubchem.ncbi.nlm.nih.gov/gene/BDNF/human>. Accessed 26 May 2024.
- . EPOR - Erythropoietin Receptor (Human). <https://pubchem.ncbi.nlm.nih.gov/gene/EPOR/human>. Accessed 19 June 2024.
- Rankinen, Tuomo, et al. “No Association between the Angiotensin-Converting Enzyme ID Polymorphism and Elite Endurance Athlete Status.” *Journal of Applied Physiology*, vol. 88, no. 5, May 2000, pp. 1571–75. journals.physiology.org (Atypon), <https://doi.org/10.1152/jappl.2000.88.5.1571>.
- Rico-Sanz, J., et al. “Associations between Cardiorespiratory Responses to Exercise and the C34T AMPD1 Gene Polymorphism in the HERITAGE Family Study.” *Physiological Genomics*, vol. 14, no. 2, July 2003, pp. 161–66. journals.physiology.org (Atypon), <https://doi.org/10.1152/physiolgenomics.00165.2002>.
- Ripa, Rashelle, et al. “Physiology, Cardiac Muscle.” *StatPearls*, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK572070/>.
- Rodríguez-Rodríguez, Diana Raquel, et al. “Genome Editing: A Perspective on the Application of CRISPR/Cas9 to Study Human Diseases (Review).” *International Journal of Molecular Medicine*, vol. 43, no. 4, Apr. 2019, pp. 1559–74. PubMed Central,

- <https://doi.org/10.3892/ijmm.2019.4112>.
- Rubio, Juan C., et al. "Frequency of the C34T Mutation of the AMPD1 Gene in World-Class Endurance Athletes: Does This Mutation Impair Performance?" *Journal of Applied Physiology*, vol. 98, no. 6, June 2005, pp. 2108–12. [journals.physiology.org \(Atypon\)](https://doi.org/10.1152/jappphysiol.01371.2004), <https://doi.org/10.1152/jappphysiol.01371.2004>.
- Santiago, Catalina, et al. "Trp64Arg Polymorphism in ADRB3 Gene Is Associated with Elite Endurance Performance." *British Journal of Sports Medicine*, vol. 45, no. 2, Feb. 2011, pp. 147–49. [bjsm.bmj.com](https://doi.org/10.1136/bjsm.2009.061366), <https://doi.org/10.1136/bjsm.2009.061366>.
- Saunders, C. J., et al. "No Association of the ACTN3 Gene R577X Polymorphism with Endurance Performance in Ironman Triathlons." *Annals of Human Genetics*, vol. 71, no. 6, 2007, pp. 777–81. Wiley Online Library, <https://doi.org/10.1111/j.1469-1809.2006.00385.x>.
- Saunders, Colleen J., et al. "The Bradykinin B2 Receptor (BDKRB2) and Endothelial Nitric Oxide Synthase 3 (NOS3) Genes and Endurance Performance during Ironman Triathlons." *Human Molecular Genetics*, vol. 15, no. 6, Mar. 2006, pp. 979–87. Silverchair, <https://doi.org/10.1093/hmg/ddl014>.
- Sawczuk, Marek, Agnieszka Maciejewska-Karlowska, et al. "Association of the ADRB2 Gly16Arg and Glu27Gln Polymorphisms with Athlete Status." *Journal of Sports Sciences*, vol. 31, no. 14, Oct. 2013, pp. 1535–44. Taylor and Francis+NEJM, <https://doi.org/10.1080/02640414.2013.786184>.
- Sawczuk, Marek, Yevgeniya I. Timshina, et al. "The -9 / 9 Polymorphism of the Bradykinin Receptor Beta 2 Gene and Athlete Status: A Study Involving Two European Cohorts." *Human Biology*, vol. 85, no. 5, Oct. 2013, pp. 741–55. bioone.org, <https://doi.org/10.3378/027.085.0511>.
- Scanavini, Daniela, et al. "Increased Frequency of the Homozygous II ACE Genotype in Italian Olympic Endurance Athletes." *European Journal of Human Genetics*, vol. 10, no. 10, Oct. 2002, pp. 576–77. [www.nature.com](https://doi.org/10.1038/sj.ejhg.5200852), <https://doi.org/10.1038/sj.ejhg.5200852>.
- Schneider, Angela J., and Theodore Friedmann. "The Problem of Doping in Sports." *Advances in Genetics*, vol. 51, Academic Press, 2006, pp. 1–9. ScienceDirect, [https://doi.org/10.1016/S0065-2660\(06\)51001-6](https://doi.org/10.1016/S0065-2660(06)51001-6).
- Schor, B., et al. "Plasma Brain-Derived Neurotrophic Factor Is Higher after Combat Training (Randori) than Incremental Ramp Test in Elite Judo Athletes." *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas E Biologicas*, vol. 52, no. 4, Apr. 2019, p. e8154. PubMed, <https://doi.org/10.1590/1414-431X20198154>.
- Scott, Robert A., et al. "No Association between Angiotensin Converting Enzyme (ACE) Gene Variation and Endurance Athlete Status in Kenyans." *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 141, no. 2, June 2005, pp. 169–75. ScienceDirect, <https://doi.org/10.1016/j.cbpb.2005.05.001>.
- Scott, Wayne, et al. "Human Skeletal Muscle Fiber Type Classifications." *Physical Therapy*, vol. 81, no. 11, Nov. 2001, pp. 1810–16. Silverchair, <https://doi.org/10.1093/ptj/81.11.1810>.
- Semenova, Ekaterina A., Elliott C. R. Hall, et al. "Genes and Athletic Performance: The 2023

- Update.” *Genes*, vol. 14, no. 6, June 2023, p. 1235. PubMed Central, <https://doi.org/10.3390/genes14061235>.
- Semenova, Ekaterina A., Hirofumi Zempo, et al. “Genome-Wide Association Study Identifies CDKN1A as a Novel Locus Associated with Muscle Fiber Composition.” *Cells*, vol. 11, no. 23, Dec. 2022, p. 3910. PubMed Central, <https://doi.org/10.3390/cells11233910>.
- Semenova, Ekaterina A., Eri Miyamoto-Mikami, et al. “The Association of HFE Gene H63D Polymorphism with Endurance Athlete Status and Aerobic Capacity: Novel Findings and a Meta-Analysis.” *European Journal of Applied Physiology*, vol. 120, no. 3, Mar. 2020, pp. 665–73. Springer Link, <https://doi.org/10.1007/s00421-020-04306-8>.
- Sensory Neuron - an Overview | ScienceDirect Topics.
<https://www.sciencedirect.com/topics/computer-science/sensory-neuron>. Accessed 24 May 2024.
- Shang, X., et al. “Association Between the ACTN3 R577X Polymorphism and Female Endurance Athletes in China.” *International Journal of Sports Medicine*, vol. 31, no. 12, Dec. 2010, pp. 913–16. www.thieme-connect.de, <https://doi.org/10.1055/s-0030-1265176>.
- Shenoy, Shweta. “Association of Angiotensin Converting Enzyme Gene Polymorphism and Indian Army Triathletes Performance.” *Asian Journal of Sports Medicine*, vol. 1, no. 3, 3, 2010, pp. 143–50. brieflands.com, <https://doi.org/10.5812/asjasm.34855>.
- Shi, Guangsen, et al. “A Rare Mutation of B1-Adrenergic Receptor Affects Sleep/Wake Behaviors.” *Neuron*, vol. 103, no. 6, Sept. 2019, pp. 1044-1055.e7. PubMed, <https://doi.org/10.1016/j.neuron.2019.07.026>.
- Singh, Khurajam Dhanachandra, and Sadashiva S. Karnik. “Angiotensin Receptors: Structure, Function, Signaling and Clinical Applications.” *Journal of Cell Signaling*, vol. 1, no. 2, June 2016, p. 111. PubMed Central, <https://doi.org/10.4172/jcs.1000111>.
- Taylor: Elite Athletes and the Gene for Angiotensin-Conve... - Google Scholar.
https://scholar.google.com/scholar_lookup?title=Elite+athletes+and+the+gene+for+angiotensin-converting+enzyme&author=Taylor,+R.R.&author=Mamotte,+C.D.&author=Fallon,+K.&author=van+Bockxmeer,+F.M.&publication_year=1999&journal=J.+Appl.+Physiol.&volume=87&pages=1035%E2%80%931037&doi=10.1152/jappl.1999.87.3.1035. Accessed 19 May 2024.
- Thakur, Sunil, et al. “Angiotensin-Converting Enzyme (ACE) Insertion/Deletion (I/D) Polymorphism as a Conjoint Regulator of Coagulation, Fibrinolytic, and RAAS Pathway in Infertility and Associated Pregnancy Complications.” *Journal of the Renin-Angiotensin-Aldosterone System: JRAAS*, vol. 2022, Nov. 2022, p. 1695769. PubMed Central, <https://doi.org/10.1155/2022/1695769>.
- The ACE I/D Polymorphism in Elite Greek Track and Field Athletes | Request PDF.
https://www.researchgate.net/publication/41089398_The_ACE_ID_polymorphism_in_elite_Greek_track_and_field_athletes. Accessed 19 May 2024.
- The Angiotensin Converting Enzyme I/D Polymorphism in Long Distance Runners - ProQuest.
<https://www.proquest.com/openview/4982fa0c707e91af301f6acbffa44ec/1?pq-origsite=gscholar&cbl=4718>. Accessed 19 May 2024.
- The Association between Ace Gene Variation and Aerobic Capacity in Winter Endurance Disciplines - PubMed. <https://pubmed.ncbi.nlm.nih.gov/24795498/>. Accessed 19 May 2024.
- Thomaes, Tom, et al. “A Genetic Predisposition Score for Muscular Endophenotypes Predicts

- the Increase in Aerobic Power after Training: The CAREGENE Study.” *BMC Genetics*, vol. 12, no. 1, Oct. 2011, p. 84. BioMed Central, <https://doi.org/10.1186/1471-2156-12-84>.
- Tigner, Alyssa, et al. “Histology, White Blood Cell.” *StatPearls*, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK563148/>.
- Tobina, Takuro, et al. “Association between the Angiotensin I-Converting Enzyme Gene Insertion/Deletion Polymorphism and Endurance Running Speed in Japanese Runners.” *The Journal of Physiological Sciences*, vol. 60, no. 5, 5, Sept. 2010, pp. 325–30. jps.biomedcentral.com, <https://doi.org/10.1007/s12576-010-0100-4>.
- Tournas, Lucille, et al. GERMLINE DOPING FOR HEIGHTENED PERFORMANCE IN SPORT.
- Trecroci, Athos, et al. “Editorial: Factors Affecting Performance and Recovery in Team Sports: A Multidimensional Perspective.” *Frontiers in Physiology*, vol. 13, Mar. 2022, p. 877879. PubMed Central, <https://doi.org/10.3389/fphys.2022.877879>.
- Truncated Erythropoietin Receptor Causes Dominantly Inherited Benign Human Erythrocytosis. - PMC. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC46538/>. Accessed 9 June 2024.
- Tsianos: Associations of Polymorphisms of Eight Muscle-or... - Google Scholar. https://scholar.google.com/scholar_lookup?title=Associations+of+polymorphisms+of+eight+muscle-+or+metabolism-related+genes+with+performance+in+Mount+Olympus+marathon+runners&author=Tsianos,+G.I.&author=Evangelou,+E.&author=Boot,+A.&author=Zillikens,+M.C.&author=van+Meurs,+J.B.&author=Uitterlinden,+A.G.&author=Ioannidis,+J.P.&publication_year=2010&journal=J.+Appl.+Physiol.&volume=108&pages=567%E2%80%93574&doi=10.1152/jappphysiol.00780.2009. Accessed 19 May 2024.
- Tsianos, Georgios, et al. “The ACE Gene Insertion/Deletion Polymorphism and Elite Endurance Swimming.” *European Journal of Applied Physiology*, vol. 92, no. 3, July 2004, pp. 360–62. Springer Link, <https://doi.org/10.1007/s00421-004-1120-7>.
- Tural, Ercan, et al. “PPAR- α and PPARGC1A Gene Variants Have Strong Effects on Aerobic Performance of Turkish Elite Endurance Athletes.” *Molecular Biology Reports*, vol. 41, no. 9, Sept. 2014, pp. 5799–804. Springer Link, <https://doi.org/10.1007/s11033-014-3453-6>.
- Turgut, Günfer, et al. “The Angiotensin Converting Enzyme I/D Polymorphism in Turkish Athletes and Sedentary Controls.” *Acta Medica (Hradec Kralove, Czech Republic)*, vol. 47, no. 2, Jan. 2019, pp. 133–36. actamedica.lfhk.cuni.cz, <https://doi.org/10.14712/18059694.2018.79>.
- Varillas-Delgado, David, et al. “Genetic Profiles to Identify Talents in Elite Endurance Athletes and Professional Football Players.” *PLOS ONE*, vol. 17, no. 9, Sept. 2022, p. e0274880. PLoS Journals, <https://doi.org/10.1371/journal.pone.0274880>.
- Vaughan, David, et al. “The Angiotensin Converting Enzyme Insertion/Deletion Polymorphism Alters the Response of Muscle Energy Supply Lines to Exercise.” *European Journal of Applied Physiology*, vol. 113, no. 7, 2013, pp. 1719–29. PubMed Central, <https://doi.org/10.1007/s00421-012-2583-6>.
- Wagoner, L. E., et al. “Polymorphisms of the Beta(2)-Adrenergic Receptor Determine Exercise Capacity in Patients with Heart Failure.” *Circulation Research*, vol. 86, no. 8, Apr. 2000, pp. 834–40. PubMed, <https://doi.org/10.1161/01.res.86.8.834>.
- Wang, Guan, et al. “Association Analysis of ACE and ACTN3 in Elite Caucasian and East Asian Swimmers.” *Medicine & Science in Sports & Exercise*, vol. 45, no. 5, May 2013, p. 892. journals.lww.com, <https://doi.org/10.1249/MSS.0b013e31827c501f>.
- Wang, Xinrong, et al. “Molecular Bases of VEGFR-2-Mediated Physiological Function and

- Pathological Role.” *Frontiers in Cell and Developmental Biology*, vol. 8, Nov. 2020. *Frontiers*, <https://doi.org/10.3389/fcell.2020.599281>.
- Watowich, Stephanie S. “The Erythropoietin Receptor: Molecular Structure and Hematopoietic Signaling Pathways.” *Journal of Investigative Medicine : The Official Publication of the American Federation for Clinical Research*, vol. 59, no. 7, Oct. 2011, pp. 1067–72. PubMed Central, <https://doi.org/10.2311/JIM.0b013e31820fb28c>.
- Wenbo, Zhang, and Zhang Yan. “The Uses of Anabolic Androgenic Steroids Among Athletes; Its Positive and Negative Aspects- A Literature Review.” *Journal of Multidisciplinary Healthcare*, vol. 16, Dec. 2023, pp. 4293–305. PubMed Central, <https://doi.org/10.2147/JMDH.S439384>.
- Williams, Alun G., et al. “Bradykinin Receptor Gene Variant and Human Physical Performance.” *Journal of Applied Physiology*, vol. 96, no. 3, Mar. 2004, pp. 938–42. journals.physiology.org (Atypon), <https://doi.org/10.1152/jappphysiol.00865.2003>.
- Williams, Olivia, and Shane R. Sergent. “Histology, Platelets.” *StatPearls*, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK557800/>.
- Wolfarth, Bernd, et al. “Association between a B2-Adrenergic Receptor Polymorphism and Elite Endurance Performance.” *Metabolism - Clinical and Experimental*, vol. 56, no. 12, Dec. 2007, pp. 1649–51. www.metabolismjournal.com, <https://doi.org/10.1016/j.metabol.2007.07.006>.
- Wong, Marty Kwok-Shing. “Subchapter 42D - Angiotensin Converting Enzyme.” *Handbook of Hormones (Second Edition)*, edited by Hironori Ando et al., Academic Press, 2021, pp. 505–08. ScienceDirect, <https://doi.org/10.1016/B978-0-12-820649-2.00128-5>.
- Yang, Nan, Daniel G. MacArthur, et al. “ACTN3 Genotype Is Associated with Human Elite Athletic Performance.” *The American Journal of Human Genetics*, vol. 73, no. 3, Sept. 2003, pp. 627–31. www.cell.com, <https://doi.org/10.1086/377590>.
- Yang, Nan, Daniel G. MacArthur, et al. “The ACTN3 R577X Polymorphism in East and West African Athletes.” *Medicine & Science in Sports & Exercise*, vol. 39, no. 11, Nov. 2007, p. 1985. journals.lww.com, <https://doi.org/10.1249/mss.0b013e31814844c9>.
- Yang, Ruoyu, et al. “Prediction and Identification of Power Performance Using Polygenic Models of Three Single-Nucleotide Polymorphisms in Chinese Elite Athletes.” *Frontiers in Genetics*, vol. 12, Oct. 2021. *Frontiers*, <https://doi.org/10.3389/fgene.2021.726552>.
- Zhang, B., et al. “The I Allele of the Angiotensin-Converting Enzyme Gene Is Associated with an Increased Percentage of Slow-Twitch Type I Fibers in Human Skeletal Muscle.” *Clinical Genetics*, vol. 63, no. 2, 2003, pp. 139–44. Wiley Online Library, <https://doi.org/10.1034/j.1399-0004.2003.00029.x>.
- Zmijewski, Piotr, Agata Grenda, et al. “Effect of BDKRB2 Gene -9/+9 Polymorphism on Training Improvements in Competitive Swimmers.” *The Journal of Strength & Conditioning Research*, vol. 30, no. 3, Mar. 2016, p. 665. journals.lww.com, <https://doi.org/10.1519/JSC.0000000000001145>.
- Zmijewski, Piotr, Paweł Ciężczyk, et al. “The NOS3 G894T (Rs1799983) and -786T/C (Rs2070744) Polymorphisms Are Associated with Elite Swimmer Status.” *Biology of Sport*, vol. 35, no. 4, 2018, pp. 313–19. www.termidia.pl, <https://doi.org/10.5114/biolsport.2018.76528>.
- Znazen, Hela, et al. “Genetic Advantageous Predisposition of Angiotensin Converting Enzyme Id Polymorphism in Tunisian Athletes.” *The Journal of Sports Medicine and Physical Fitness*, vol. 56, no. 6, June 2016, pp. 724–30.

Decoding Fragile X Syndrome: Genetic and Epigenetic Modifications of FMR1 Gene and Therapeutic Implications By Mai Lam

Abstract

Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability and autism spectrum disorder, primarily resulting from a trinucleotide CGG repeat expansion in the *FMR1* gene. This paper explores the role of genetic and epigenetic modifications of *FMR1* in the pathogenesis and clinical manifestations of FXS. Specifically, the review investigates how the methylation and silencing of the *FMR1* gene lead to the absence of fragile X mental retardation protein (FMRP), which is crucial for normal synaptic function. Furthermore, it discusses the interplay between genetic mutations and epigenetic mechanisms, such as DNA methylation, histone modifications, and chromatin remodeling, in driving disease severity. In addition to summarizing the molecular basis of FXS, this paper also examines therapeutic approaches targeting these genetic and epigenetic alterations. These include emerging gene-editing tools like CRISPR/Cas systems, antisense oligonucleotides, and pharmacological interventions aimed at reversing abnormal methylation patterns or restoring FMRP expression. Despite significant progress, challenges remain in translating these findings into effective treatments. The review also highlights future directions and the need for further research to refine these therapeutic strategies and develop personalized interventions for FXS.

Introduction

Fragile X syndrome (FXS) is a genetic disorder that results from a mutation in the *FMR1* gene located on the X chromosome (Hunter et al., 2024). It is recognized as the second most prevalent cause of intellectual disability (ID) after trisomy 21 and the leading cause of inherited ID (Ben-Yosef et al., 2019). The exact number of people who have FXS is unknown, but it is estimated that approximately 1 in 7,000 males and 1 in 11,000 females are affected globally (Hunter et al., 2014). Notably, males are typically more severely affected than females due to the presence of a single X chromosome, which exacerbates the impact of the dysfunctional gene.

The fragile X mental retardation protein (FMRP), encoded by the *FMR1* gene, is the fundamental component of FXS. FMRP is a critical RNA-binding protein responsible for transporting mRNA, including its own, out of the nucleus and repressing translation (Darnell et al., 2011). This regulatory function of FMRP is important for synaptic development and plasticity, as it controls the production and translation of other proteins necessary for synapse formation and maintenance (Sidorov et al., 2013). Synapses are essential for effective communication between neurons, and their proper function is vital for normal cognitive development and neural activity.

FXS occurs when the *FMR1* gene on the X chromosome has a trinucleotide repeat expansion of CGG repeats, with affected individuals having over 200 repeats compared to the normal 5 to 44 (Maddalena et al., 2001). This expansion leads to the reduction or elimination of FMRP production, affecting synaptic function and contributing to FXS symptoms. Males with a full mutation of FXS always exhibit symptoms, including significant intellectual disability and

delayed motor and speech development (Hagerman, 2002). They have distinct facial features such as a large head, long face, prominent forehead and chin, and protruding ears. They may also have behavioral and connective tissue problems like hyperactivity and social anxiety or double-jointed fingers and hyperflexible joints. Many are diagnosed with autism spectrum disorder, showing poor eye contact, repetitive speech, impulse control issues, and distractibility. Physical problems can include strabismus, mitral valve prolapse, aortic root dilation, and occasional dermatological issues. Females with the full *FMRI* mutation can also exhibit these features, though approximately 50% are intellectually disabled, usually less severely than males (Hunter et al., 2024).

Other than genetic alterations, epigenetic modifications also play a significant role in the pathogenesis of FXS. Epigenetic changes, such as DNA methylation, histone modification, and RNA-associated silencing, can influence gene expression without altering the underlying DNA sequence. In the context of FXS, hypermethylation of the *FMRI* gene's promoter region is a critical epigenetic modification leading to the gene's transcriptional silencing (Nobile et al., 2021)

Addressing this silencing through gene-targeting medicines offer potential therapeutic approaches. One notable strategy is the use of antisense oligonucleotides (ASOs) that target the mRNA of the *FMRI* gene to restore normal protein synthesis. Studies have demonstrated that these ASOs can effectively reduce the levels of abnormal mRNA and, consequently, the pathological effects associated with the *FMRI* gene mutation (Shah et al., 2023). Additionally, CRISPR-Cas9 gene-editing technology is being explored to directly correct the genetic mutation responsible for FXS. This cutting-edge approach aims to permanently alter the DNA sequence, offering a potential long-term solution for individuals affected by the syndrome (Yrigollen et al., 2019).

This review examines the role of genetic and epigenetic modifications of the *FMRI* gene in the development and severity of Fragile X syndrome (FXS). Additionally, it summarizes current and emerging therapeutic strategies aimed at targeting these modifications.

History of Fragile X Syndrome

Fragile X Syndrome was first described in 1943 by British physician J.P. Martin and his colleague Julia Bell, who observed a pattern of intellectual disability in families that seemed to be linked to the X chromosome (Martin et al., 1943). This initial observation led to the condition being referred to as Martin-Bell syndrome. However, it was not until 1969 that Herbert Lubs discovered a "marker X chromosome" in affected individuals, characterized by a fragile site at the end of the long arm of the X chromosome (Lubs, 1969). This discovery provided the basis for the term Fragile X Syndrome.

The breakthrough in understanding the genetic basis of FXS came in 1991 when the *FMRI* gene on the X chromosome was identified as the locus responsible for the condition (Verkerk et al., 1991). They discovered that a mutation in this gene, specifically an expansion of the CGG trinucleotide repeat, leads to the inactivation of the gene, resulting in the absence or

deficiency of the fragile X mental retardation protein (FMRP), which is crucial for normal neural development.

Diagnostic techniques for FXS have evolved significantly over time. Early methods involved cytogenetic analyses that relied on identifying the fragile site on the X chromosome under specific laboratory conditions (Sutherland, 1977). These methods were not only labor-intensive but also lacked sensitivity, especially in detecting mosaic forms of the mutation. The introduction of Southern blot analysis in the 1980s provided a more reliable means of diagnosing FXS by measuring the size of the CGG repeat expansion (Oberlé et al., 1991).

In the following decades, polymerase chain reaction (PCR)-based protocols were developed, allowing for faster and more precise detection of the *FMRI* mutation (Hagerman et al., 2002). The most recent advancement in diagnostic technology is triplet-primed PCR (TP-PCR), which enhances the detection of expanded alleles, even in mosaic forms, thus improving diagnostic accuracy and efficiency (Chen et al., 2011).

Genetic alterations of Fragile X Syndrome

1. FMR1 Gene and CGG Repeat Expansion

The *FMRI* gene, located on the Xq27.3 region of the X chromosome, spans approximately 38 kilobases (kb) of genomic DNA and consists of 17 exons. The most significant part of the *FMRI* gene is its 5' untranslated region (UTR), where the CGG trinucleotide repeat sequence is located. This region is crucial because its length determines the gene's activity and the likelihood of FXS (Loesch et al., 2003). Normally, the CGG repeat number in the *FMRI* gene ranges from 5 to 44 repeats, which does not affect gene function or protein production (Santoro et al., 2012).

Individuals with 55 to 200 CGG repeats in the *FMRI* gene are classified as premutation carriers. These repeats are unstable and can expand during maternal transmission, significantly increasing the likelihood of the offspring inheriting a full mutation. The full mutation is characterized by an insertion of more than 200 CGG repeats, leading to hypermethylation of the *FMRI* gene's promoter region. This hypermethylation effectively silences the gene, preventing the synthesis of fragile X mental retardation protein (FMRP), which is essential for normal synaptic function and plasticity (Jacquemont et al., 2011; Coffee et al., 1999). This disruption in FMRP production results in the characteristic neurological and developmental impairments observed in individuals with Fragile X Syndrome (FXS).

The expansion of CGG repeats in the *FMRI* gene occurs primarily through a mechanism known as slipped-strand mispairing during DNA replication. This process can form secondary structures, such as hairpins, which facilitate the addition of extra repeats, which is known as an insertion. This instability is particularly pronounced during germ cell development, especially in oogenesis, where the CGG repeats are more prone to expansion. As a result, the length of the repeat can increase significantly from one generation to the next, further exacerbating the risk of transmitting a full mutation to offspring (Usdin, 2008).

2. Loss of FMRP Function

Fragile X mental retardation protein is a critical RNA-binding protein that plays a significant role in the transport, stability, and translation of specific mRNAs within neurons (Darnell et al., 2011). It binds to polyribosomes, the cellular machinery responsible for protein synthesis, and regulates the production of proteins essential for synaptic structure and function. This regulatory function is vital for maintaining proper neuronal communication and plasticity, which are crucial for learning and memory (Santoro et al., 2012). Without FMRP, the precise control of protein synthesis is disrupted, leading to various neuronal and cognitive deficits observed in Fragile X Syndrome (FXS).

The absence of FMRP, primarily due to the mutation of *FMRI* gene, leads to significant synaptic dysfunction, characterized by the dysregulation of synaptic protein synthesis. This dysregulation results in abnormal synaptic structure and impaired synaptic plasticity, which are directly associated with the cognitive deficits and behavioral abnormalities seen in individuals with FXS (Bagni et al., 2012). One of the hallmark features of FXS is the presence of abnormal dendritic spine morphology. Neurons in patients with FXS exhibit an increased number of long, thin dendritic spines, indicative of immature synapses. These morphological abnormalities correlate with impaired synaptic signaling and connectivity, further contributing to the neurological symptoms of the disorder (Santoro et al., 2012).

The loss of FMRP affects several critical molecular pathways. One major pathway is the metabotropic glutamate receptor 5 (mGluR5) pathway. In the absence of FMRP, there is enhanced mGluR5 signaling, leading to exaggerated long-term depression (LTD), which contributes to synaptic dysfunction (Dölen et al., 2007). Another important pathway affected by the loss of FMRP is the phosphoinositide 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) pathway. FMRP interacts with this pathway to regulate protein synthesis. Dysregulation of the PI3K/mTOR pathway due to the absence of FMRP results in altered synaptic protein synthesis and cellular signaling, further exacerbating the neurological impairments in FXS (Hoeffler et al., 2012). Understanding these pathways is crucial for developing targeted therapeutic strategies to mitigate the effects of FMRP deficiency in FXS.

3. Other Health Implications of the *FMRI* Gene and FMRP

Beyond FXS, the *FMRI* gene and its protein product, FMRP, have broader implications in other health conditions. Premutation carriers of the *FMRI* gene, with 55 to 200 CGG repeats, are at risk for developing Fragile X-associated tremor/ataxia syndrome (FXTAS) and Fragile X-associated primary ovarian insufficiency (FXPOI) (Oostra et al., 2009). These conditions further highlight the importance of understanding the *FMRI* gene's regulation and the wide-ranging impact of its dysregulation. FXTAS is a neurodegenerative disorder affecting primarily older adult males, characterized by intention tremor, gait ataxia, parkinsonism, and cognitive decline (Jacquemont et al., 2003). FXPOI affects approximately 20% of female premutation carriers and is characterized by ovarian dysfunction, leading to infertility and early menopause (Sullivan et al., 2005). Emerging evidence also suggests that *FMRI* gene

dysregulation may play roles in other psychiatric and neurodevelopmental disorders, even in individuals without CGG repeat expansions (Sabaratnam et al., 2001). These conditions further underscore the importance of understanding the *FMR1* gene's regulation and the wide-ranging impact of its dysregulation.

Epigenetic Modifications in Fragile X Syndrome

1. DNA Methylation

Epigenetic modifications, particularly DNA methylation, are central to the development of FXS. In individuals with a full mutation of the *FMR1* gene, an expansion of CGG repeats leads to hypermethylation of cytosine residues in the CpG islands of the *FMR1* promoter, silencing the gene by preventing the binding of essential transcription factors (Nobile et al., 2021). This gene silencing results in the absence of FMRP, the protein whose deficiency causes the symptoms of FXS.

Moreover, the epigenetic landscape in FXS can vary among individuals, with some exhibiting mosaicism—a condition where both methylated and unmethylated alleles of the *FMR1* gene coexist. This mosaicism can lead to a milder phenotype in FXS, as the presence of unmethylated alleles allows for partial expression of the *FMR1* gene and some production of FMRP. This partial gene expression can mitigate the severity of symptoms, highlighting the complex interplay between epigenetic modifications and clinical outcomes in FXS (Kumari & Usdin, 2009; Jacquemont et al., 2007).

To assess *FMR1* methylation, traditional techniques like Southern blotting have been used, providing an overview of methylation status and CGG repeat size (Coffee et al., 1999). More advanced methods like Methylation-Specific PCR (MS-PCR) allow for detailed analysis by distinguishing between methylated and unmethylated sequences in bisulfite-treated DNA (Kumari & Usdin, 2009). Next-Generation Sequencing (NGS) further enhances methylation analysis, offering comprehensive insights into specific loci, including the *FMR1* gene (Kinde et al., 2011; Laird, 2010). These advanced techniques enhance our understanding of the epigenetic mechanisms in FXS and hold potential for improving diagnosis and treatment.

2. Histone Modifications

Histone proteins, around which DNA is wrapped to form nucleosomes, undergo various post-translational modifications such as acetylation, methylation, phosphorylation, and ubiquitination. These modifications constitute the "histone code," which influences chromatin structure and gene expression (Santoro et al., 2012; Jenuwein & Allis, 2001).

In FXS, the histones associated with the *FMR1* promoter region exhibit reduced acetylation, correlating with a repressive chromatin state. Histone deacetylases (HDACs) remove acetyl groups from histones, contributing to chromatin compaction and gene silencing (Sutcliffe et al., 1992). Specific methylation marks, such as H3K9me3 and H3K27me3, are enriched at the *FMR1* locus in FXS patients. These repressive marks inhibit transcriptional activity by recruiting

chromatin remodeling complexes that maintain a heterochromatic state (Jacquemont et al., 2011).

Research is ongoing to explore epigenetic therapies targeting histone modifications. Histone deacetylase inhibitors (HDACi) are being investigated for their potential to reactivate *FMR1* expression by promoting a more open chromatin configuration (Hagerman et al., 2009).

3. RNA-Mediated Silencing

Non-coding RNAs (ncRNAs), including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), play crucial roles in gene regulation. These RNAs can modulate gene expression through various mechanisms such as RNA interference (RNAi) and chromatin remodeling. RNAi is a biological process where RNA molecules inhibit gene expression or translation by neutralizing targeted mRNA molecules (Fire et al., 1998). Chromatin remodeling involves the dynamic modification of chromatin architecture to allow access of condensed genomic DNA to the regulatory transcription machinery proteins, and it is crucial for regulating gene expression (Kornberg & Lorch, 1999; Zhou et al., 2019).

MicroRNAs such as miR-9 and miR-124 have been implicated in the regulation of *FMR1* expression. They can bind to the 3' UTR of *FMR1* mRNA, influencing its stability and translation. Dysregulation of these miRNAs may contribute to the altered gene expression observed in FXS (Zhou et al., 2019). Some ncRNAs transcribed from the *FMR1* locus may interfere with its transcription, further contributing to gene silencing.

RNA-induced silencing complexes (RISC) and RNA-induced transcriptional silencing (RITS) complexes are involved in RNAi mechanisms. These complexes can guide the methylation of DNA and histones, reinforcing the silencing of the *FMR1* gene (Zhou et al., 2019)

Therapeutic Approaches

1. Pharmacological Treatments

a. Targeting mGluR5 Pathway

Preclinical studies and early clinical trials have investigated mGluR5 antagonists, such as mavoglurant (AFQ056) and basimglurant (RG7090). These drugs aim to reduce exaggerated mGluR5 signaling in FXS (Youssef et al., 2018). However, clinical trials have yielded mixed results, with some showing improvement in behavior and cognition, while others did not demonstrate significant benefits (Jacquemont et al., 2011).

b. GABAergic Modulation

Drugs like arbaclofen, a GABA-B receptor agonist, have been tested to enhance inhibitory signaling and reduce hyperexcitability in FXS patients (Berry-Kravis et al., 2017). Clinical trials have shown some promise, particularly in reducing social anxiety and improving adaptive behavior (Newman et al., 2022).

c. mTOR Pathway Inhibitors

Rapamycin and Analogues: In preclinical models, rapamycin and its analogues have shown potential in normalizing synaptic protein synthesis and improving behavioral deficits. It has been difficult to apply these findings to human therapy due to side effects and the requirement for ongoing treatment. (Hoeffler et al., 2012).

2. Gene Therapy

Gene Reactivation Strategies

The CRISPR/Cas9 genome-editing technology has been used to target the methylated CGG repeats in the *FMR1* gene. By editing the DNA sequence, CRISPR/Cas9 can potentially reactivate the silenced *FMR1* gene, allowing for the production of FMRP. Early research has shown that this approach can reduce methylation and partially restore *FMR1* expression in cellular models (Liu et al., 2016). However, challenges remain in delivering CRISPR/Cas9 efficiently and safely to the brain, as well as ensuring precise and controlled editing without off-target effects.

Gaps in Research

Despite significant advances in understanding FXS, several gaps remain in the current research that must be addressed to develop more effective treatments and improve patient outcomes.

1. Understanding Molecular Mechanisms

While the role of CGG repeat expansion and FMRP deficiency in FXS is well-established, the exact molecular mechanisms by which these changes lead to neurological and behavioral symptoms are still not fully understood. Further research is needed to explain the pathways and interactions involved in the development of FXS. For instance, the downstream effects of FMRP absence on synaptic function, plasticity, and neural circuitry require more detailed investigation (Bear et al., 2004). Additionally, understanding how other genetic and epigenetic factors modulate *FMR1* expression and contribute to phenotypic variability in FXS can provide insights into more precise therapeutic targets (Bagni et al., 2012).

2. Development of Effective Therapies

Current therapies for FXS, including pharmacological treatments and gene therapy, have shown promise but also face significant challenges. There is a critical need for more research to identify effective treatment targets, improve drug delivery systems, and enhance the safety and efficacy of therapeutic interventions. Specifically, the development of mGluR5 antagonists and GABAergic modulators has shown mixed results, highlighting the necessity for optimization in terms of dosage, timing, and patient selection (Youssef et al., 2018). Furthermore, gene therapy approaches, such as CRISPR/Cas9, require advancements in delivery methods to target the brain

efficiently and safely, minimizing off-target effects and ensuring long-term therapeutic benefits (Liu et al., 2016).

3. Environmental and Lifestyle Factors

The influence of environmental and lifestyle factors on the severity and progression of FXS is not well understood. Factors such as diet, physical activity, social interactions, and educational interventions may play significant roles in modulating the symptoms of FXS. Research exploring these aspects could lead to non-pharmacological strategies that complement medical treatments, improving overall patient outcomes. Investigating the impact of early educational and behavioral interventions could provide evidence for the development of comprehensive care plans that integrate medical, educational, and psychosocial support (Bailey et al., 2008).

4. Animal Models

While animal models have provided valuable insights into FXS, they do not fully recapitulate the human condition. Developing more accurate and diverse models, including non-human primates and patient-derived induced pluripotent stem cells (iPSCs), can enhance our understanding of FXS pathophysiology and facilitate the testing of new therapies (Kaufmann et al., 2017). These models can be used to study the effects of genetic and epigenetic modifications, as well as to screen for potential drug candidates and evaluate their efficacy and safety in a controlled environment.

Limitations and Future Directions

However, this exploration is limited by several factors. One key limitation is the reliance on predominantly preclinical studies, such as those involving animal models or cell cultures. While these models provide valuable insights, they do not fully replicate the human condition, leading to potential gaps in translating findings into effective clinical therapies. Additionally, the review's focus on well-established pathways like mGluR5 and CRISPR/Cas9 can overlook emerging, less-explored molecular targets that could hold therapeutic potential. The fast-paced development in the fields of gene editing and epigenetics means that some of the review's cited studies may become outdated as newer findings emerge.

Looking forward, future research directions for this literature review could involve a more targeted analysis of recent advancements in personalized medicine approaches for FXS. As more is understood about the variability in FXS phenotypes, there is a growing need to explore therapies tailored to individual genetic and epigenetic profiles (Elhawary et al., 2023). Further studies could also focus on optimizing drug delivery systems, especially for gene therapies like CRISPR/Cas9, to ensure their safe and effective application in humans (Huang et al., 2022). Moreover, expanding the literature review to include more recent clinical trials involving diverse populations could enhance the applicability of findings across different demographic groups (Berry-Kravis et al., 2013). Finally, incorporating more interdisciplinary research—combining

insights from genetics, psychology, and education—could provide a broader understanding of how environmental factors and early interventions influence the progression of FXS and its treatment outcomes (Scott et al., 2020; Nfxf, 2021).

Conclusion

In conclusion, the genetic and epigenetic modifications of the *FMRI* gene play a crucial role in Fragile X Syndrome and related conditions. Epigenetic modifications, such as the hypermethylation of the *FMRI* gene, directly cause FXS by silencing the production of essential FMRP proteins. While significant progress has been made in understanding these mechanisms and developing therapeutic strategies, many challenges and gaps remain. Continued research is important to fully explain the molecular basis of FXS, develop personalized treatment approaches, and ensure the long-term efficacy and safety of therapies. Moreover, exploring the broader implications of *FMRI* gene dysregulation can provide valuable insights into a range of health conditions, ultimately improving the diagnosis and treatment of individuals affected by these disorders.

Works Cited

- Bagni, Claudia et al. "Fragile X Syndrome: Causes, Diagnosis, Mechanisms, and Therapeutics." *The Journal of Clinical Investigation*, vol. 122, no. 12, 2012, pp. 4314-22. doi:10.1172/JCI63141.
- Bailey, Donald B Jr et al. "Co-occurring Conditions Associated with FMR1 Gene Variations: Findings from a National Parent Survey." *American Journal of Medical Genetics. Part A*, vol. 146A, no. 16, 2008, pp. 2060-9. doi:10.1002/ajmg.a.32439.
- Bear, Mark F et al. "The mGluR Theory of Fragile X Mental Retardation." *Trends in Neurosciences*, vol. 27, no. 7, 2004, pp. 370-7. doi:10.1016/j.tins.2004.04.009.
- Ben-Yosef, Dalit, and Yoav Mayshar. *Fragile-X Syndrome: Methods and Protocols*. Humana Press, 2019.
- Berry-Kravis, Elizabeth et al. "Arbaclofen in Fragile X Syndrome: Results of Phase 3 Trials." *Journal of Neurodevelopmental Disorders*, vol. 9, 2017, doi:10.1186/s11689-016-9181-6.
- Berry-Kravis, Elizabeth et al. "Outcome Measures for Clinical Trials in Fragile X Syndrome." *Journal of Developmental and Behavioral Pediatrics: JDBP*, vol. 34, no. 7, 2013, pp. 508-22. doi:10.1097/DBP.0b013e31829d1f20.
- Chen, L., et al. "High-Resolution Methylation Polymerase Chain Reaction for Fragile X Analysis: Evidence for Novel FMR1 Methylation Patterns Undetected in Southern Blot Analyses." *Genetics in Medicine*, vol. 13, no. 6, 2011, pp. 528-38. doi:10.1097/GIM.0b013e31820a780f.
- Coffee, B et al. "Acetylated Histones Are Associated with FMR1 in Normal but Not Fragile X-Syndrome Cells." *Nature Genetics*, vol. 22, no. 1, 1999, pp. 98-101. doi:10.1038/8807.
- Elhawary, N.A., AlJahdali, I.A., Abumansour, I.S. et al. "Phenotypic Variability to Medication Management: An Update on Fragile X Syndrome." *Human Genomics*, vol. 17, 2023. <https://doi.org/10.1186/s40246-023-00507-2>.
- Fire, A et al. "Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis Elegans*." *Nature*, vol. 391, no. 6669, 1998, pp. 806-11. doi:10.1038/35888.
- Hagerman, R. J., and P. J. Hagerman. *Fragile X Syndrome*. In P. Howlin & O. Udwin (Eds.), *Outcomes in Neurodevelopmental and Genetic Disorders*, Cambridge University Press, 2002, pp. 198-219. <https://doi.org/10.1017/CBO9780511543876.009>.
- Hagerman, Randi J et al. "Advances in the Treatment of Fragile X Syndrome." *Pediatrics*, vol. 123, no. 1, 2009, pp. 378-90. doi:10.1542/peds.2008-0317.
- Hoeffler, C A et al. "Altered mTOR Signaling and Enhanced CYFIP2 Expression Levels in Subjects with Fragile X Syndrome." *Genes, Brain, and Behavior*, vol. 11, no. 3, 2012, pp. 332-41. doi:10.1111/j.1601-183X.2012.00768.
- Huang, Jie et al. "CRISPR/Cas Systems: Delivery and Application in Gene Therapy." *Frontiers in Bioengineering and Biotechnology*, vol. 10, 2022. doi:10.3389/fbioe.2022.942325.
- Hunter, J., et al. "Epidemiology of Fragile X Syndrome: A Systematic Review and Meta-Analysis." *American Journal of Medical Genetics A.*, vol. 164A, no. 7, 2014, pp. 1648-58.

- Hunter, Jessica Ezzell. "FMR1 Disorders." GeneReviews [Internet]., U.S. National Library of Medicine, 16 May 2024, www.ncbi.nlm.nih.gov/books/NBK1384/.
- Jacquemont, Sébastien et al. "Epigenetic Modification of the FMR1 Gene in Fragile X Syndrome Is Associated with Differential Response to the mGluR5 Antagonist AFQ056." *Science Translational Medicine*, vol. 3, no. 64, 2011. doi:10.1126/scitranslmed.3001708.
- Jacquemont, Sébastien et al. "Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates." *American Journal of Human Genetics*, vol. 72, no. 4, 2003, pp. 869-78. doi:10.1086/374321.
- Jacquemont, Sebastien et al. "Fragile-X Syndrome and Fragile X-Associated Tremor/Ataxia Syndrome: Two Faces of FMR1." *The Lancet. Neurology*, vol. 6, no. 1, 2007, pp. 45-55. doi:10.1016/S1474-4422(06)70676-7.
- J.C. Darnell et al. "FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism." *Cell.*, vol. 146, 2011, pp. 247-61.
- Jenuwein, T, and C D Allis. "Translating the Histone Code." *Science (New York, N.Y.)*, vol. 293, no. 5532, 2001, pp. 1074-80. doi:10.1126/science.1063127.
- Kaufmann, Walter E et al. "Autism Spectrum Disorder in Fragile X Syndrome: Cooccurring Conditions and Current Treatment." *Pediatrics*, vol. 139, Suppl. 3, 2017, pp. S194-S206. doi:10.1542/peds.2016-1159F.
- Kinde, Isaac et al. "Detection and Quantification of Rare Mutations with Massively Parallel Sequencing." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 23, 2011, pp. 9530-5. doi:10.1073/pnas.1105422108.
- Kornberg, R D, and Y Lorch. "Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome." *Cell*, vol. 98, no. 3, 1999, pp. 285-94. doi:10.1016/s0092-8674(00)81958-3.
- Kumari, Daman, and Karen Usdin. "Chromatin Remodeling in the Noncoding Repeat Expansion Diseases." *The Journal of Biological Chemistry*, vol. 284, no. 12, 2009, pp. 7413-7. doi:10.1074/jbc.R800026200.
- Laird, P. "Principles and Challenges of Genome-Wide DNA Methylation Analysis." *Nature Reviews Genetics*, vol. 11, 2010, pp. 191-203. <https://doi.org/10.1038/nrg2732>.
- Liu, X Shawn et al. "Editing DNA Methylation in the Mammalian Genome." *Cell*, vol. 167, no. 1, 2016, pp. 233-47.e17. doi:10.1016/j.cell.2016.08.056.
- Loesch, Danuta Z et al. "Effect of the Fragile X Status Categories and the Fragile X Mental Retardation Protein Levels on Executive Functioning in Males and Females with Fragile X." *Neuropsychology*, vol. 17, no. 4, 2003, pp. 646-57. doi:10.1037/0894-4105.17.4.646.
- Lubs, H A. "A Marker X Chromosome." *American Journal of Human Genetics*, vol. 21, no. 3, 1969, PP. 231-44
- Maddalena, A et al. "Technical Standards and Guidelines for Fragile X: The First of a Series of Disease-Specific Supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics." *Genetics in Medicine*, vol. 3, no. 3, 2001, pp. 200-5. doi:10.1097/00125817-200105000-00008.
- Martin, J P, and J Bell. "A PEDIGREE OF MENTAL DEFECT SHOWING SEX-LINKAGE." *Journal of neurology and psychiatry* vol. 6,3-4 (1943): 154-7. doi:10.1136/jnnp.6.3-4.154

- Newman, Michelle G et al. "Psychotherapeutic treatments for generalized anxiety disorder: cognitive and behavioral therapies, enhancement strategies, and emerging efforts." *Expert review of neurotherapeutics* vol. 22,9 (2022): 751-770. doi:10.1080/14737175.2022.2125800
- Nfxf. "Early Intervention for Infants and Toddlers with Fragile X Syndrome: NFXF." National Fragile X Foundation, 23 June 2021, fragilex.org/treatment-and-intervention/early-intervention-infants-toddlers-fragile-x-syndrome/.
- Nobile, Veronica et al. "DNA Methylation, Mechanisms of FMR1 Inactivation and Therapeutic Perspectives for Fragile X Syndrome." *Biomolecules* vol. 11,2 296. 16 Feb. 2021, doi:10.3390/biom11020296
- Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*. 1991 May 24;252(5009):1097-102. doi: 10.1126/science.252.5009.1097. PMID: 2031184.
- Oostra, Ben A, and Rob Willemsen. "FMR1: a gene with three faces." *Biochimica et biophysica acta* vol. 1790,6 (2009): 467-77. doi:10.1016/j.bbagen.2009.02.007
- Sabaratham, M et al. "Epilepsy and EEG findings in 18 males with fragile X syndrome." *Seizure* vol. 10,1 (2001): 60-3. doi:10.1053/seiz.2000.0492
- Santoro, Michael R et al. "Molecular mechanisms of fragile X syndrome: a twenty-year perspective." *Annual review of pathology* vol. 7 (2012): 219-45. doi:10.1146/annurev-pathol-011811-132457
- Scott, Haleigh, et al. "Cognitive Training Deep Dive: The Impact of Child, Training Behavior and Environmental Factors within a Controlled Trial of Cogmed for Fragile X Syndrome." MDPI, Multidisciplinary Digital Publishing Institute, 25 Sept. 2020, www.mdpi.com/2076-3425/10/10/671.
- Shah, Sneha et al. "Antisense oligonucleotide rescue of CGG expansion-dependent FMR1 mis-splicing in fragile X syndrome restores FMRP." *Proceedings of the National Academy of Sciences of the United States of America* vol. 120,27 (2023): e2302534120. doi:10.1073/pnas.2302534120
- Sidorov, Michael S et al. "Fragile X mental retardation protein and synaptic plasticity." *Molecular brain* vol. 6 15. 8 Apr. 2013, doi:10.1186/1756-6606-6-15
- Sullivan, A K et al. "Association of FMR1 repeat size with ovarian dysfunction." *Human reproduction (Oxford, England)* vol. 20,2 (2005): 402-12. doi:10.1093/humrep/deh635
- Sutcliffe, J S et al. "DNA methylation represses FMR-1 transcription in fragile X syndrome." *Human molecular genetics* vol. 1,6 (1992): 397-400. doi:10.1093/hmg/1.6.397
- Sutherland GR. Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. *Science*. 1977 Jul 15;197(4300):265-6. doi: 10.1126/science.877551. PMID: 877551.
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991 May 31;65(5):905-14. doi: 10.1016/0092-8674(91)90397-h. PMID: 1710175.

- Youssef, Eriene A et al. "Effect of the mGluR5-NAM Basimglurant on Behavior in Adolescents and Adults with Fragile X Syndrome in a Randomized, Double-Blind, Placebo-Controlled Trial: FragXis Phase 2 Results." *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* vol. 43,3 (2018): 503-512. doi:10.1038/npp.2017.177
- Yrigollen, Carolyn M, and Beverly L Davidson. "CRISPR to the Rescue: Advances in Gene Editing for the FMR1 Gene." *Brain sciences* vol. 9,1 17. 21 Jan. 2019, doi:10.3390/brainsci9010017
- Zhou, Yafang et al. "Non-coding RNA in Fragile X Syndrome and Converging Mechanisms Shared by Related Disorders." *Frontiers in genetics* vol. 10 139. 1 Mar. 2019, doi:10.3389/fgene.2019.00139

A Comparative Analysis of the Effect of Liquidity on the Price of Bitcoin

By Rushil Jaiswal

Abstract

This paper delves into the intricate relationship between liquidity indicators and the price dynamics of Bitcoin, a prominent cryptocurrency. Liquidity, a fundamental aspect of financial markets, profoundly influences market stability and efficiency. Leveraging statistical analysis and AI modeling techniques, my study explores various liquidity metrics—including trading volume, bid-ask spread, order book depth, volatility, number of transactions, bid and ask sums as separate indicators, and market depth—to assess their impact on the price of Bitcoin. The findings offer valuable insights into the factors driving Bitcoin price movements and shed light on the role of liquidity in cryptocurrency markets. Through correlation analysis as well as three different machine learning models: random forests, XGBoost, and linear regression, my study evaluates the significance of individual liquidity factors and their relationships with Bitcoin prices. The best performing models were the random forest and XGBoost regressors where I identified that the volatility was the feature that was the most informative of the model's performance. My research contributes to advancing our understanding of liquidity and price discovery in cryptocurrency markets and underscores the need for future studies to explore alternative factors and mechanisms shaping cryptocurrency prices. By embracing the findings and continuously refining analytical approaches, researchers can navigate the evolving landscape of cryptocurrency trading, ultimately enhancing market efficiency and informing regulatory decisions.

Keywords: Bitcoin, Liquidity Indicators, Trading Volume, Bid-Ask Spread, Order Book Depth, Market Depth, Price Dynamics, Cryptocurrency Markets.

I. Introduction

In the dynamic realm of cryptocurrency markets, the interaction between liquidity and price movements is a pivotal focus for researchers, traders, and investors alike. As a developing and widely traded digital asset, Bitcoin has garnered increasing attention for its inherent volatility, characterized by abrupt price fluctuations.

Bitcoin (BTC), developed by Satoshi Nakamoto in 2009, is a decentralized cryptocurrency underpinned by blockchain technology (Investopedia). Unlike purchasing stock in a company, which grants ownership in the company itself, buying Bitcoin provides ownership of the cryptocurrency based on the amount purchased. Bitcoin's price is primarily influenced by supply and market demand; however, there are additional factors that are often overlooked. The volatility and market liquidity of Bitcoin, as well as other liquidity indicators such as trading volume and bid-ask spreads, also play significant roles in determining its price dynamics. This research aims to explore these lesser-considered factors to gain a more comprehensive understanding of the forces driving Bitcoin's market value.

Understanding the relationship between liquidity and price changes is essential for grasping the intricacies of the cryptocurrency market and developing informed strategies for navigating its complexities. This paper embarks on a comprehensive comparative analysis, aiming to dissect and evaluate the influence of liquidity on the price dynamics of bitcoin. Liquidity, a multifaceted concept, encompasses the ease with which an asset can be bought or sold in the market without causing significant price disruptions (Bitcoin.com, Investopedia). This research addresses two fundamental questions surrounding the liquidity-pricing relationship in the Bitcoin market: [1] Do liquidity metrics accurately capture the market dynamics? [2] Are there discernible patterns in the market impact of large trades concerning liquidity fluctuations?. This comparative approach allows us to juxtapose different market conditions, timeframes, or liquidity rules, offering a nuanced understanding of how liquidity dynamics may vary and impact Bitcoin's price behavior. To carry out these comparisons, correlation analysis as well as three types of machine learning models – random forests, XGBoost, and linear regression – are applied. The models with the best performance were the Random Forest and XGBoost models with $R^2 = 0.91$. Furthermore, I also applied feature analysis to understand the most important factors driving the models' predictions. Here, we find that for the highest performing model, the feature that is most influential is the volatility.

This research is pivotal in advancing our understanding of how liquidity metrics impact Bitcoin's price dynamics. While much of the existing literature has focused on traditional financial assets, this study highlights how specific liquidity indicators- such as bid-ask spread, trading volume, and volatility, bid prices, and ask prices - affect Bitcoin, a unique, and highly volatile cryptocurrency. By integrating high frequency and low frequency liquidity measures, the study fills a critical gap in the literature by providing a more nuanced analysis of how these factors influence Bitcoin price. The findings offer valuable insights for investors and traders seeking to navigate the complexities of cryptocurrency markets, and they can also guide policymakers in developing informed regulations that address the unique challenges and risks associated with digital currencies.

Cryptocurrency research has surged, with many studies exploring the intricate relationship between Bitcoin price dynamics and liquidity. Due to Bitcoin's volatility, decentralized nature, and global reach, understanding liquidity's impact on price fluctuations is crucial.

Early studies by Kyle (1985) and Amihud et al. (1986) laid the groundwork for understanding liquidity through bid-ask spreads and trading volume. Theoretical frameworks from market microstructure studies by O'Hara (1995) and Madhavan (2000) further dissected liquidity and price dynamics. Recent research by Yelowitz (2015) examines the market microstructure of cryptocurrency exchanges, highlighting nuances in liquidity provision and price impact.

Abdi and Rinaldo (2017) found low-frequency transactions-based liquidity measures to be effective compared to high-frequency benchmarks. Their study highlights the performance of various liquidity estimators, including the Corwin and Schultz (2012) estimator, in describing

time-series variations across different observation frequencies, trading venues, and cryptocurrencies.

Building on this literature, this paper analyzes the impact of liquidity on Bitcoin's price changes, comparing high-frequency liquidity measures with low-frequency measures for Bitcoin (BTC). This investigation reveals that low-frequency liquidity measures provide accurate liquidity estimates in cryptocurrency markets incorporating new empirical evidence.

II. Methods

A. Data

The data for analysis was sourced from data.bitcoinity.org. It encompasses price data and changes in Bitcoin price over the past five years, along with metrics pertaining to the selected liquidity indicators for the same duration. This dataset was deemed suitable for the research, providing a comprehensive foundation for examining the relationship between liquidity factors and Bitcoin prices over time. This research employs a meticulous methodology and leverages relevant data sources to illuminate the intricate interplay between liquidity dynamics and Bitcoin price movements, offering valuable insights for market participants and researchers alike.

With respect to issues with the data, missing data was addressed by removing data for all the other features, making sure we only keep the data for the parts where all the information is there. In addition, the price (target of the prediction) was log-transformed. We found that transforming the data in this manner yielded better empirical results and amplified the signal present in the data.

B. Feature Selection

After a review of the literature, a set of liquidity indicators was chosen to proceed with the analysis. These indicators were selected based on their perceived ability to effectively reflect the liquidity dynamics within the Bitcoin market. The chosen liquidity indicators include:

Bid-Ask Spread: This metric signifies the difference between the highest price a buyer is willing to pay (bid) and the lowest price a seller is willing to accept (ask). A narrower spread typically indicates higher liquidity.

Bid-Ask Sum: This metric aggregates the total volume of bids and asks within the market. It provides insight into the overall depth of the market and the level of trading activity.

Volatility: Volatility measures the degree of variation in Bitcoin's price over a certain period. High volatility may indicate lower liquidity as it can deter market participants due to increased risk.

Trades Per Minute: This metric quantifies the frequency of trades occurring within a minute. A higher number of trades per minute may suggest higher liquidity as it reflects active participation in the market.

Trading Volume: Trading volume represents the total number of Bitcoin units traded within a specified period. Higher trading volume generally indicates higher liquidity.

Number of Transactions: This metric counts the total number of transactions executed on the Bitcoin network within a given timeframe. It offers insights into the activity level and participation in the Bitcoin ecosystem.

C. Statistical and Machine Learning Methods

The main graphs I used to analyze my results compared the liquidity factors to the price of Bitcoin to understand their correlation. Lastly, I created AI models using XGBoost, Linear regression, and Random Forest Regression and calculated the R-squared values for each model to evaluate their performance.

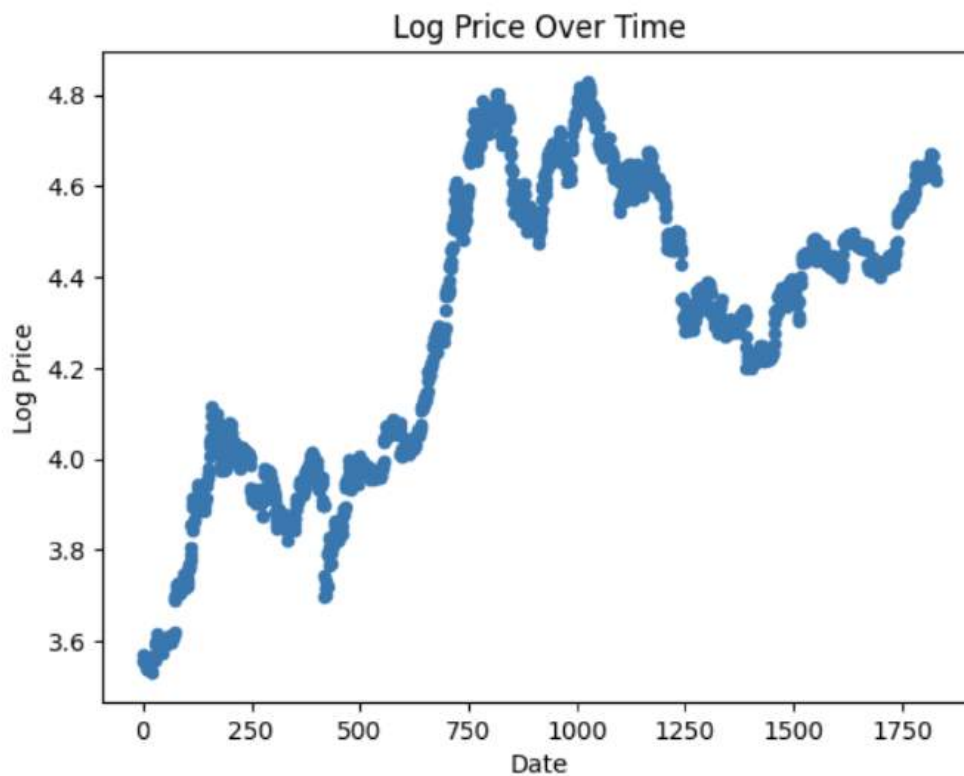


Figure 1:

The log price change of Bitcoin over the years from around 2019 to present date.



Figure 2:
The log volatility change of Bitcoin affected its price over the years, from around 2019 to the present date.

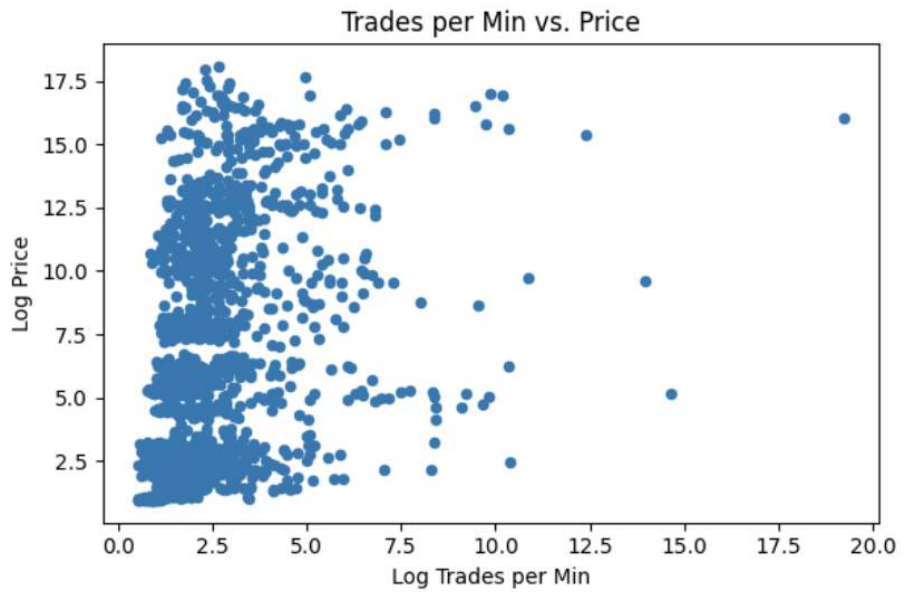


Figure 3:
The log transactions change of Bitcoin affect on the price of Bitcoin over the years from around 2019 to the present date.

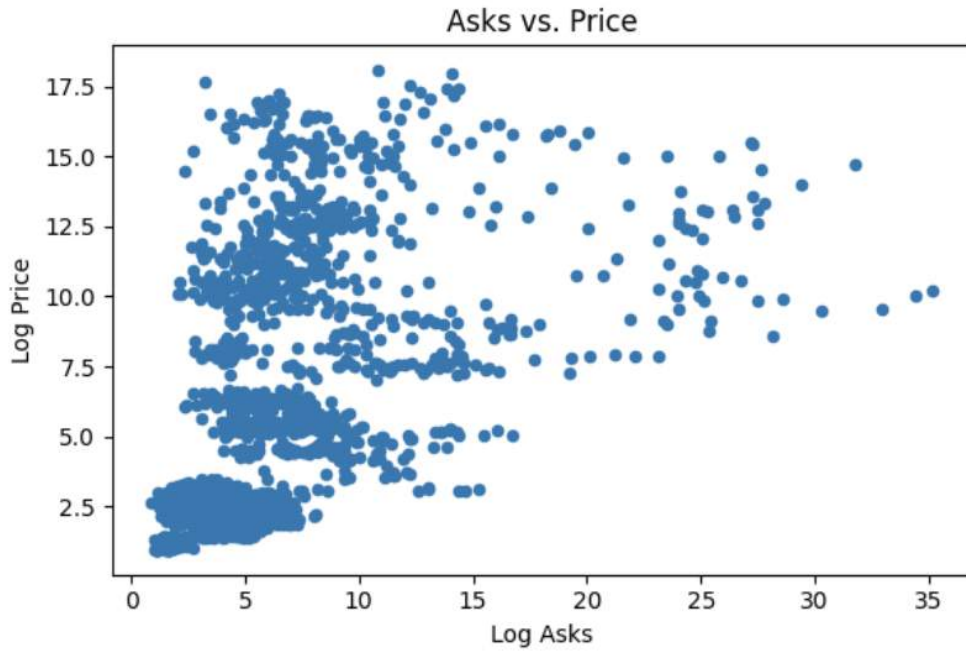


Figure 4:
The log asks change of Bitcoin affects the price of Bitcoin over the years from around 2019 to the present date.

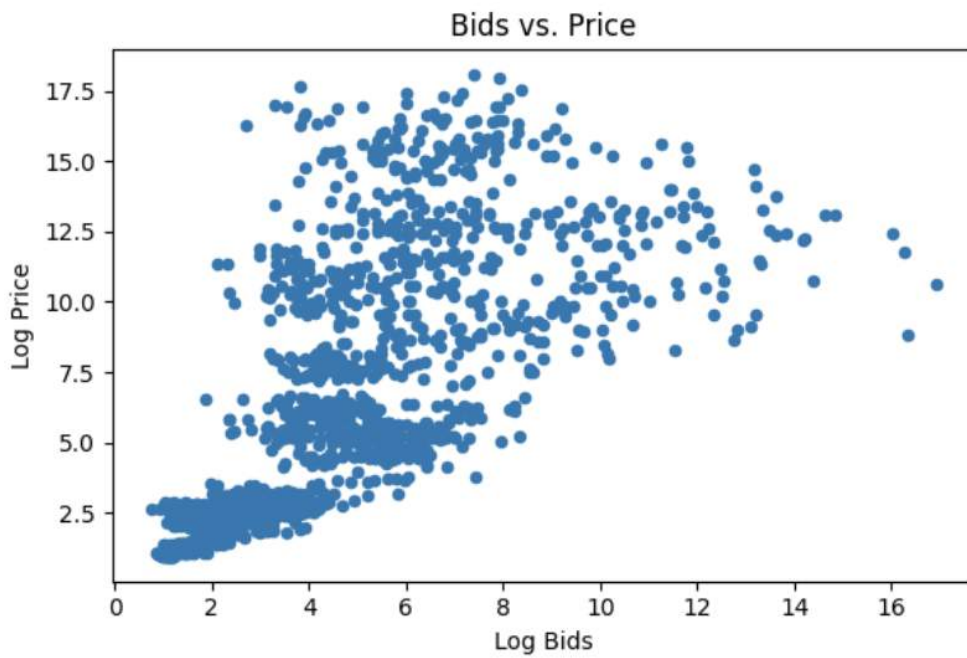


Figure 5:
The log bids change of Bitcoin affects the price of Bitcoin over the years from around 2019 to the present date.

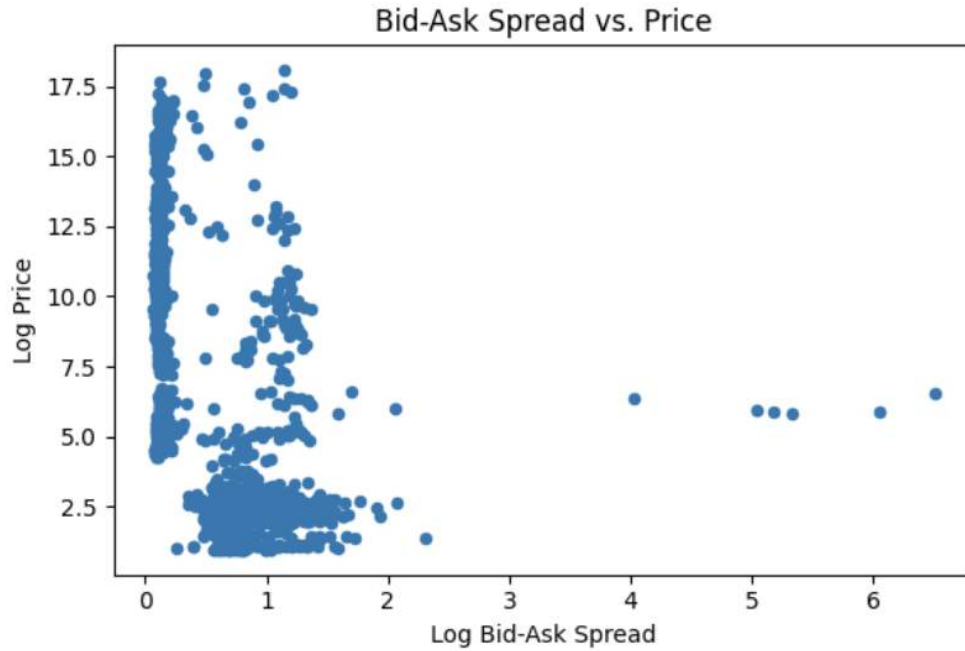


Figure 6:
The Log Bid-Ask Spread change of Bitcoin affects the price of Bitcoin over the years from around 2019 to the present date.

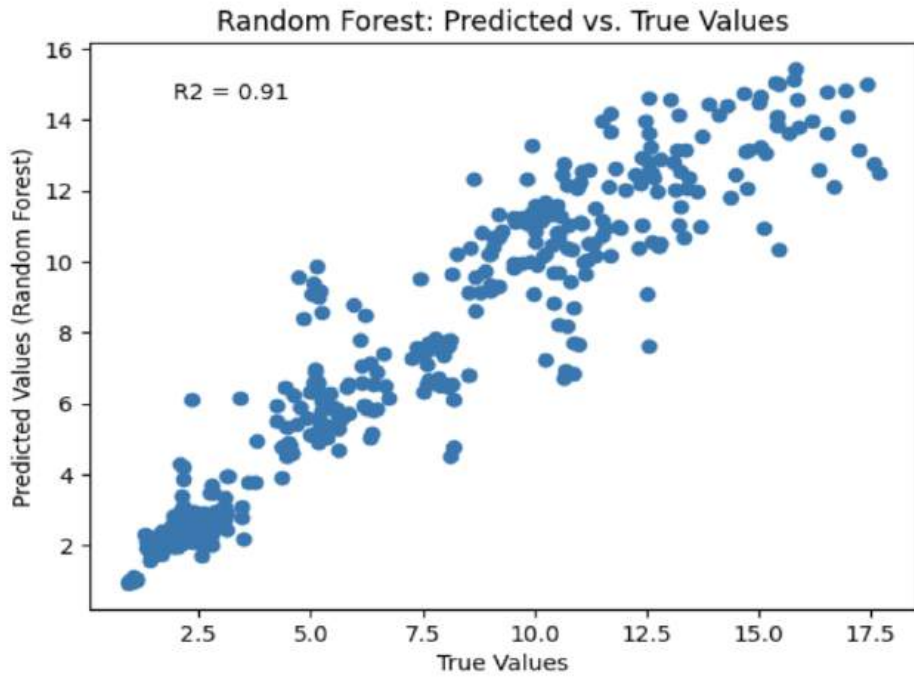


Figure 7:
The Predicted and True R2 Values for the Random Forest Model

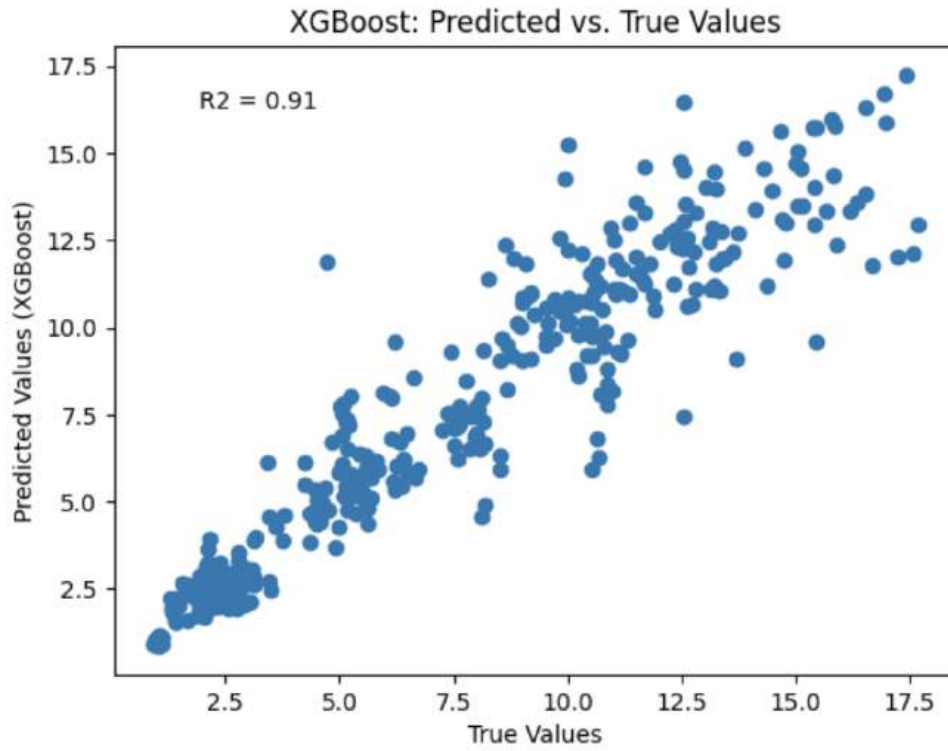


Figure 8:
The Predicted and True R2 Values for the XGBoost Model

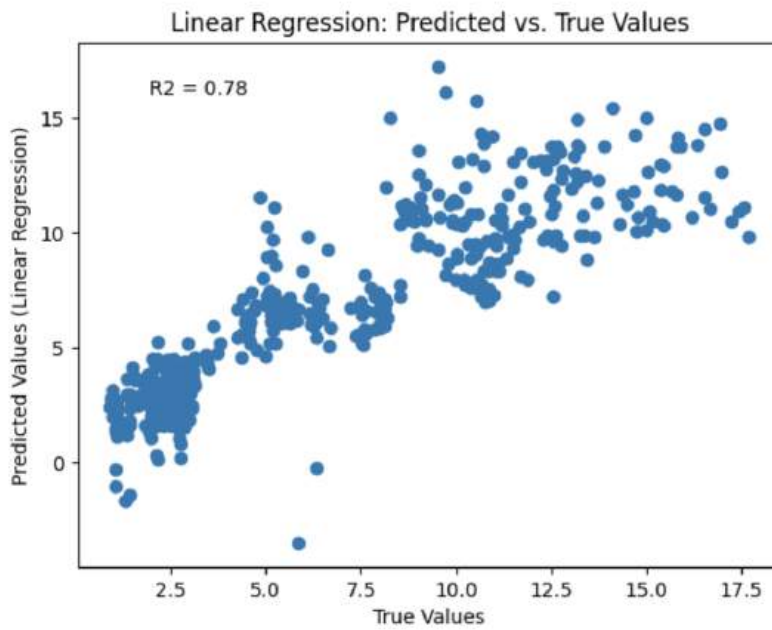


Figure 9:
The Predicted and True R2 Values for the Linear Regression Model

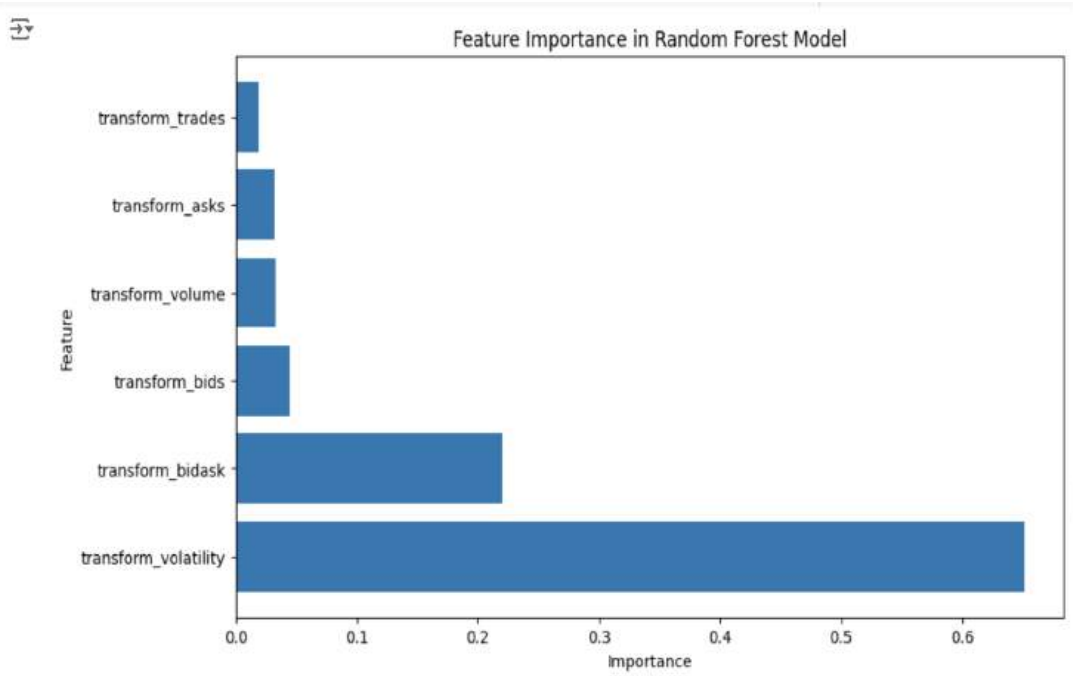


Figure 10:
Feature Importance in the Random Forest Model

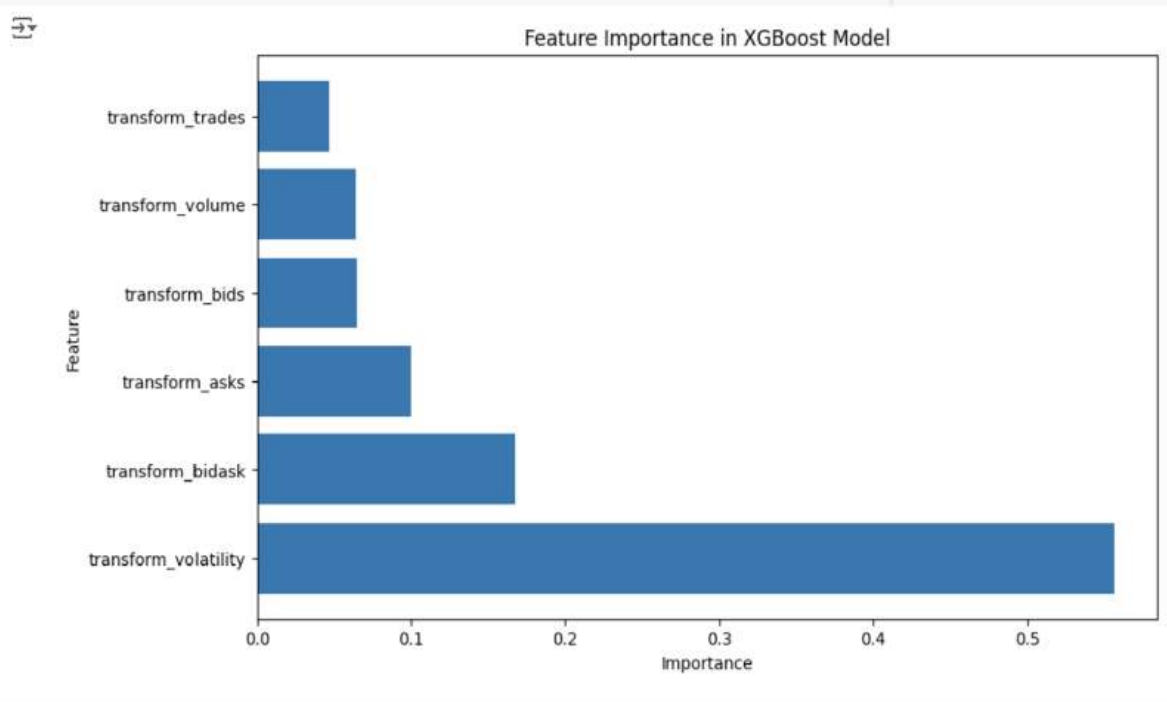


Figure 11:
Feature Importance in XGBoost Model

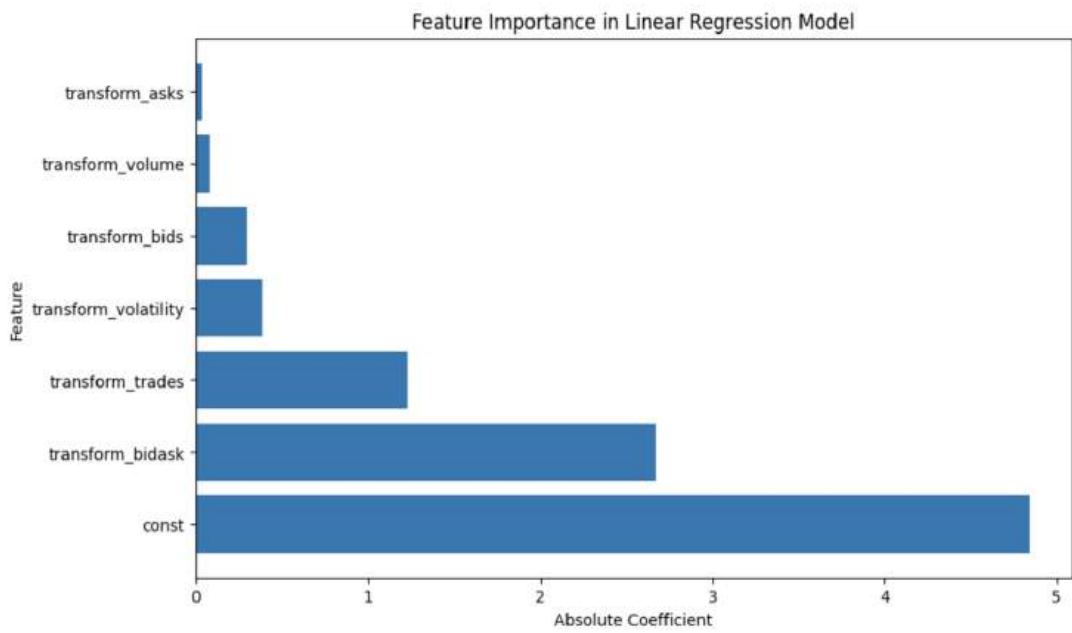


Figure 12:
Feature Importance in Linear Regression

III. Results

This section presents the findings of the comparative analysis of liquidity metrics and their impact on Bitcoin price dynamics. The analysis involves correlation studies and the application of machine learning models, specifically Random Forest, XGBoost, and Linear Regression, to evaluate the predictive power of various liquidity indicators. The log-transformed price of Bitcoin on Figure 1 over the past five years shows the significant fluctuations and overall upward trend in its value. This provides a baseline for understanding how various liquidity metrics correlate with Bitcoin's price movements. The next few figures provide visual insights into how liquidity indicators have evolved and potentially influenced Bitcoin prices. Figure 2 with the Log Volatility shows the fluctuations in Bitcoin's volatility, indicating periods of high and low market stability. Figure 3 with the Log number of transactions illustrates the frequency of Bitcoin transactions over time, reflecting market activity. Figure 4 with the Log Asks price captures the minimum amount someone is willing to accept for something in the market. In Figure 5, Log Bids captures the maximum someone is ready to offer for something in the market. Figure 6 with the Log Bid-Ask Spread shows the difference between the highest bid and the lowest ask prices, indicating market liquidity. A correlation analysis was conducted to examine the relationships between Bitcoin prices and the selected liquidity metrics. The results indicate significant correlations between Bitcoin prices and various liquidity factors, highlighting their potential impact on price movements. The highest correlation was found between volatility and Bitcoin prices, as volatility is identified as the most significant predictor of Bitcoin prices in both the Random Forest and XGBoost models, with more than 60% importance. In contrast, the number of trades exhibited the lowest correlation, contributing less than 5% to the models' predictive power.

I used three machine learning models to predict Bitcoin prices based on liquidity metrics: Random Forest Regressor, XGBoost, and Linear Regression. The performance of each model was evaluated using R-squared values.

Model	R squared score
XGBRegressor	0.91
Linear Regression	0.76
Random Forest Regressor	0.91

Table 1:

This table shows the R-squared score for the XGBRegressor, Linear Regression, and Random Forest Regressor Models.

The Random Forest Regressor and XGBoost Regressor demonstrated the highest predictive accuracy with an R-squared value of 0.91, followed by Linear Regression. Figure 7

shows the scatter plot of predicted vs. true values for the Random Forest model. It shows a strong alignment, indicating high predictive accuracy. The scatter plot for the XGBoost in the Figure 8 model also shows good alignment, with an R-squared value of 0.91. The scatter plot in Figure 9 for the Linear Regression model shows a lower alignment compared to the other models, with an R-squared value of 0.76. The feature importance plot for the Random Forest model in Figure 10 indicates that volatility is the most significant predictor of Bitcoin prices, followed by other liquidity metrics. The XGBoost model in Figure 11 also highlights volatility as a key feature, consistent with the Random Forest model. The Linear Regression model in Figure 12 identifies bid-ask spread as the most influential factor, differing from the other models.

The results underscore the critical role of liquidity metrics in predicting Bitcoin prices. Volatility emerged as the most informative feature in both the Random Forest and XGBoost models, highlighting its significant influence on Bitcoin price dynamics. The high predictive accuracy of the Random Forest model and XGBoost model, with R-squared values of 0.91, suggests that incorporating liquidity indicators can substantially enhance the understanding and forecasting of Bitcoin price movements.

Despite efforts to conduct a thorough analysis, several limitations must be acknowledged in this research, which may impact the interpretation and generalization of the findings:

Data Quality and Availability: The analysis heavily relies on data availability and quality from the external source data.bitcoinity.org. While efforts were made to ensure the reliability of the data, inaccuracies or inconsistencies within the dataset could introduce biases or limitations to the analysis.

Selection of Liquidity Indicators: The choice of liquidity indicators is subjective and may not capture the full spectrum of liquidity dynamics within the Bitcoin market. Alternative indicators or additional dimensions of liquidity not included in this study could yield different insights into the relationship between liquidity and Bitcoin prices.

Correlation vs. Causation: The analysis primarily identifies correlations between liquidity factors and Bitcoin prices. However, establishing causality between liquidity dynamics and price movements is challenging due to market dynamics' complex and multifaceted nature. Other unobserved variables or external factors may influence liquidity and Bitcoin prices, confounding the interpretation of results.

Market Conditions and Dynamics: The analysis spans a specific period, and market conditions and dynamics during this period may not be representative of all market environments. Changes in regulatory policies, technological advancements, macroeconomic factors, or market sentiment could influence liquidity dynamics and Bitcoin prices differently over time.

Sample Size and Timeframe: The analysis encompasses a five-year timeframe, which may limit the generalizability of findings to other time periods. Additionally, the dataset's sample size may not capture all relevant variations and patterns in liquidity and Bitcoin prices, particularly during extreme market volatility or structural shifts.

Model Complexity and Simplifications: The analysis may employ simplified models or methodologies to explore the relationship between liquidity and Bitcoin prices. While these simplifications facilitate interpretation and analysis, they may oversimplify the underlying mechanisms governing liquidity dynamics and price movements, leading to potential biases or inaccuracies.

IV. Conclusion

In conclusion, this study offers a foundational understanding of the intricate relationship between liquidity and Bitcoin price dynamics. While the analysis reveals the correlation between liquidity indicators and Bitcoin prices, it also opens avenues for future research to delve deeper into this complex relationship. Future studies could explore additional liquidity metrics beyond those considered in this analysis, such as depth of order book, market depth, and transaction cost metrics, to gain a more comprehensive understanding of liquidity dynamics in cryptocurrency markets. Moreover, employing advanced analytical techniques, including machine learning algorithms and time-series modeling, could provide further insights into the nuanced interactions between liquidity and price movements. Additionally, considering broader market and behavioral factors, such as investor sentiment, regulatory developments, and macroeconomic indicators, would enrich our understanding of the drivers of liquidity and price in cryptocurrency markets. By addressing these research gaps, future studies can contribute to developing more robust trading strategies, enhance market efficiency, and provide valuable insights for regulatory decision-making in the rapidly evolving cryptocurrency landscape.

Works Cited

- Bajpai, Prableen. "Liquidity of Bitcoin." Edited by Erika Rasure. Investopedia, www.investopedia.com/articles/investing/112914/liquidity-bitcoins.asp. Accessed 7 June 2024.
- Brauneis, Alexander, et al. "How to measure the liquidity of cryptocurrency markets?" *Journal of Banking & Finance*, vol. 1124, 2021, www.sciencedirect.com/science/article/pii/S0378426620303022. Accessed 7 June 2024.
- "data.bitcoinity.org (beta version)." Bitcoinity.org, data.bitcoinity.org/markets/price/30d/USD?c=e&t=l. Accessed 7 June 2024.
- Mansa, Julius. "What Is Bitcoin? How to Mine, Buy, and Use It." Edited by Amanda Jackson. Investopedia, www.investopedia.com/terms/b/bitcoin.asp. Accessed 7 June 2024.
- "What is Bitcoin?" Bitcoin.com, www.bitcoin.com/get-started/what-is-bitcoin/. Accessed 7 June 2024.

Estimating the Speed and Decay Rate of Muons By Aritro Ganguly

Abstract

Cosmic Rays, beams of particles from outer space, bombard the Earth's atmosphere incessantly. These rays are made up of subatomic particles, such as muons, leptons, and fermions, energy ranges from 1 GeV to 10^8 TeV. In this study, we focus on one such subatomic particle- Muons. An apparatus was designed and constructed using circuitry to detect muons eventually reaching the Earth's surface. Specifically, scintillators and photomultipliers were used to count muons, and the recorded data was analyzed to determine their speed and decay rate. The apparatus efficiently determined the muons counts in a fixed time interval. To ensure the theoretical underpinnings of the data, aspects of special relativity were applied. Finally, we discuss a potential application of this study in mapping the unknown contours of the ocean floor.

Introduction

Cosmic Rays are beams of particles from outer space. They bombard our atmosphere and contain an array of sub-atomic particles such as muons, leptons, and fermions. Muons may be perceived as insignificant particles but have exceptional use. Muons decay rapidly, which the experiment will determine, making controlling them very difficult [1].

Detecting Muons

Detecting microscopic particles has been an endeavor for years, but the sub-atomic size is a major complication. In my experiment, I utilized an apparatus to magnify and detect muons [2] *Physical Components*. To properly detect muons, the electrical component consists of three parts: (1)The Scintillator and Photomultiplier; (2) the Circuitry; and (3) the Software.

The Scintillators and Photomultiplier are two components that work symbiotically to allow muons to be detected efficiently. The scintillators are constructed of elements that will enable parts of radiation [muons] to be seen with light flashes [3]. These flashes occur when the muons decay creating a burst of energy. The flashes will be registered and added to the count. Following that, photomultipliers are crucial to the success of scintillators.

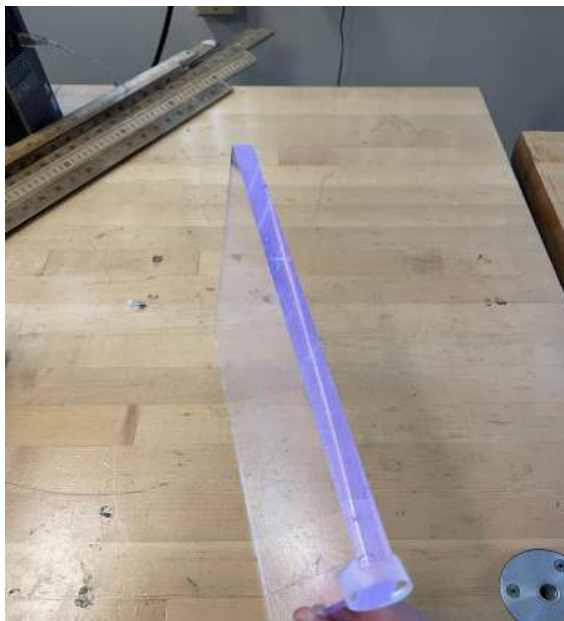


Fig 1. Scintillator utilized in the experiment

Photomultipliers receive the flashes of light from the scintillator and send signals to the electrical components. Photomultipliers multiply the strength of the light by producing a photoelectron which then multiplies the electron at the second dynode repeating with more electrons till the final dynode [3]. As a result, the flashes of light are multiplied and easier to detect.

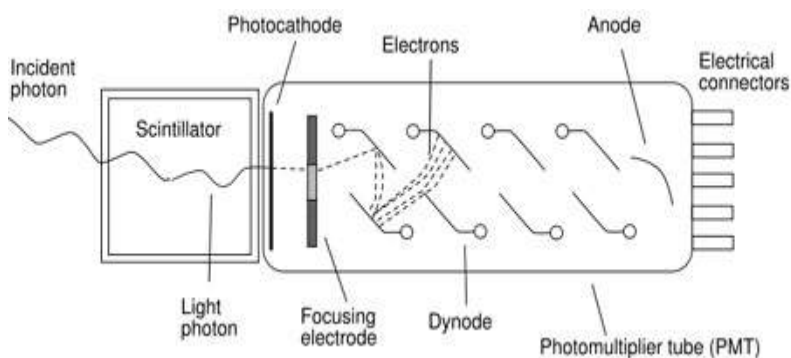


Fig2.Photomultiplier diagram [3]



Fig 3. Photomultiplier used in the experiment

In the experiment, four plastic scintillators were used to capture a field of muons in a half-foot by half-foot configuration. One photomultiplier was attached to each scintillator to expand the power of the plastic scintillators. The photomultipliers used in the experiments were constructed of a photocathode and photomultiplier Tube (PMT) [4]. The photomultipliers are about 6 inches in length. With a photocathode and photomultiplier attached, the high-energy

muons can pass through the scintillator, the photomultipliers, and eventually the electrical components.

Electrical Component. The electrical component plays a crucial role in registering the raw data. These components consist of a set of circuit boards with a screen attached that counts the muons in a specified time limit. Simultaneously with the count, a flashing blue light is seen on the board when a muon is detected. The flashing light is pointed directly from the scintillator to the photomultiplier. The eight wires connect to the scintillators/photomultipliers and provide power to the system.



Fig 4. Circuit used to detect muon counts

Software. The software is the final step in translating the collected data for analysis. The software used in the experiment is RUCosmicproduct, a Rutgers University [2]. RUCosmic records muons and their lifetimes within a specified time limit in microseconds or nanoseconds. This collected data was interpreted by graphing results. RUCosmic also allowed the data to be transported to the graphing software Logger Pro where the official results were derived.

Analyzing Data to Determine Muon Speed and Decay Rate

Four trials were conducted in 6-microsecond timeframes. To ensure completeness of the dataset, each trial had one more scintillator than the prior one starting from one to four. Such a construct calculated the muon speed more accurately, with the increased scintillators elongating the region for the muon to decay.

time microsecond	counts	Error	count1	error1	count2	error2	count3	error3
0.2	265	16.279	38	6.164	35	5.916	192	13.856
0.4	67	8.185	31	5.568	25	5.000	11	3.317
0.6	63	7.937	23	4.796	28	5.292	12	3.464
0.8	43	6.557	18	4.243	14	3.742	11	3.317
1	47	6.856	13	3.606	22	4.690	12	3.464
1.2	34	5.831	16	4.000	12	3.464	6	2.449
1.4	42	6.481	20	4.472	14	3.742	8	2.828
1.6	35	5.916	12	3.464	17	4.123	6	2.449
1.8	38	6.164	12	3.464	14	3.742	12	3.464
2	42	6.481	13	3.606	18	4.243	11	3.317
2.2	28	5.292	6	2.449	10	3.162	12	3.464
2.4	31	5.568	10	3.162	14	3.742	7	2.646
2.6	22	4.690	8	2.828	9	3.000	5	2.236
2.8	28	5.292	9	2.908	11	3.606	6	2.449
3	22	4.690	10	3.162	11	3.317	4	2.000
3.2	19	4.359	8	2.828	8	2.828	3	1.732
3.4	17	4.123	7	2.646	7	2.646	2	1.414
3.6	22	4.690	7	2.646	4	2.000	11	3.317
3.8	12	3.464	6	2.449	3	1.732	4	2.000
4	6	2.449	3	1.732	3	1.732	2	1.414
4.2	4	2.000	4	2.000	0	0.000	0	0.000
4.4	4	2.000	4	2.000	0	0.000	0	0.000
4.6	2	1.414	2	1.414	0	0.000	0	0.000
4.8	1	1.000	1	1.000	0	0.000	0	0.000
5	1	1.000	1	1.000	0	0.000	0	0.000
5.2	1	1.000	1	1.000	0	0.000	0	0.000
5.4	0	0.000	0	0.000	0	0.000	0	0.000
5.6	0	0.000	0	0.000	0	0.000	0	0.000
5.8	0	0.000	0	0.000	0	0.000	0	0.000
6	0	0.000	0	0.000	0	0.000	0	0.000

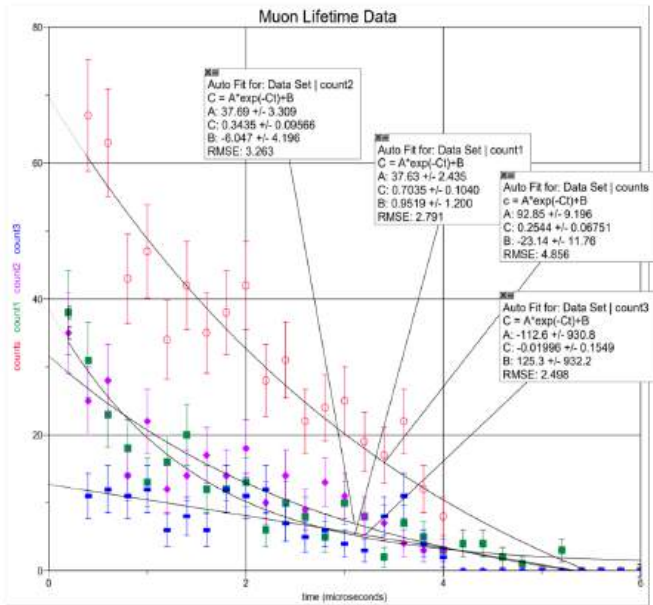


Fig 5. Graph of the muon count for each trial

Data Set 1

Time (ms)	counts0	error0	counts 1	error1	counts2	error2	counts 3	error3
0.2	265	16.279	38	6.164	35	5.916	192	13.856
0.4	67	8.185	31	5.568	25	5.000	11	3.317
0.6	63	7.937	23	4.796	28	5.292	12	3.464
0.8	43	6.557	18	4.243	14	3.742	11	3.317
1	47	6.856	13	3.606	22	4.690	12	3.464
1.2	34	5.831	16	4.000	12	3.464	6	2.449
1.4	42	6.481	20	4.472	14	3.742	8	2.828
1.6	35	5.916	12	3.464	17	4.123	6	2.449
1.8	38	6.164	12	3.464	14	3.742	12	3.464
2	42	6.481	13	3.606	18	4.243	11	3.317
2.2	28	5.292	6	2.449	10	3.162	12	3.464
2.4	31	5.568	10	3.162	14	3.742	7	2.646
2.6	22	4.690	8	2.828	9	3.000	5	2.236

2.8	24	4.899	5	2.236	13	3.606	6	2.449
3	25	5.000	10	3.162	11	3.317	4	2.000
3.2	19	4.359	8	2.828	8	2.828	3	1.732
3.4	17	4.123	2	1.414	7	2.646	8	2.828
3.6	22	4.690	7	2.646	4	2.000	11	3.317
3.8	12	3.464	5	2.236	3	1.732	4	2.000
4	8	2.828	3	1.732	3	1.732	2	1.414
4.2	4	2.000	4	2.000	0	0.000	0	0.000
4.4	4	2.000	4	2.000	0	0.000	0	0.000
4.6	2	1.414	2	1.414	0	0.000	0	0.000
4.8	1	1.000	1	1.000	0	0.000	0	0.000
5	0	0.000	0	0.000	0	0.000	0	0.000
5.2	3	1.732	3	1.732	0	0.000	0	0.000
5.4	0	0.000	0	0.000	0	0.000	0	0.000
5.6	0	0.000	0	0.000	0	0.000	0	0.000
5.8	0	0.000	0	0.000	0	0.000	0	0.000
6	0	0.000	0	0.000	0	0.000	0	0.000

Fig 6. Table of each trial count and the error

Decaying of Tau Leptons releases muons. When a Tau Lepton decays, many particles are released, but one such particle is the muons. In the graph above, we interpreted the decaying of tau particles by observing a sudden rise in muon counts [5].

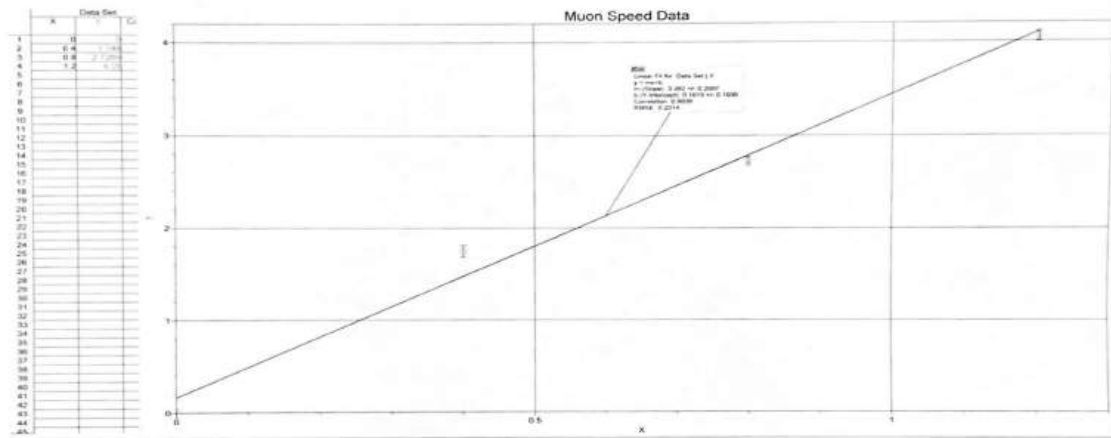


Fig 7. Graph to determine speed of the muon

Data Set 2

X	Y
0	0
0.4	1.746
0.8	2.726
1.2	4.050

Fig 8. Table for speed calculation

To determine the speed of the muons, a simple physics principle of kinematics is utilized ($t = x/v$). Using $y=mx + b$, a linear interpretation of the speed would be the slope with the time on the y-axis and the distance on the x-axis. The x-axis is the distance traveled by the muons between successive plates is about 0.4m. The y-axis corresponds to the time it takes from the muon to travel between the plates. The slope of the above best-fit line is 3.282 ± 0.259 . In the final step to calculate the speed of the muons, we take the reciprocal of the slope and transform it to m/s from m/nanoseconds. This results in the v_{calc} of approximately $(3.047 \pm 0.240) \times 10^9$. This simplifies to the speed of $(1.016 \pm 0.08)c$, where c is the speed of light. The estimated speed of muons is about $0.936c$.

What This Tells Us

Muons form about 15 km from the Earth's surface [6]. The estimated velocity is about 280,605,741 m/s. Muons have a known half-life of about $1.525 \mu\text{s}$ and a mean lifetime of about $2.2 \mu\text{s}$ [7]. That means that muons would take about $50.3366341 \mu\text{s}$ to travel 15km. However, based on the lifetime equation $I = e^{-t/\tau}$ [7], $1.15669026 \times 10^{-10} \%$ of muons would only survive and reach the Earth's surface.

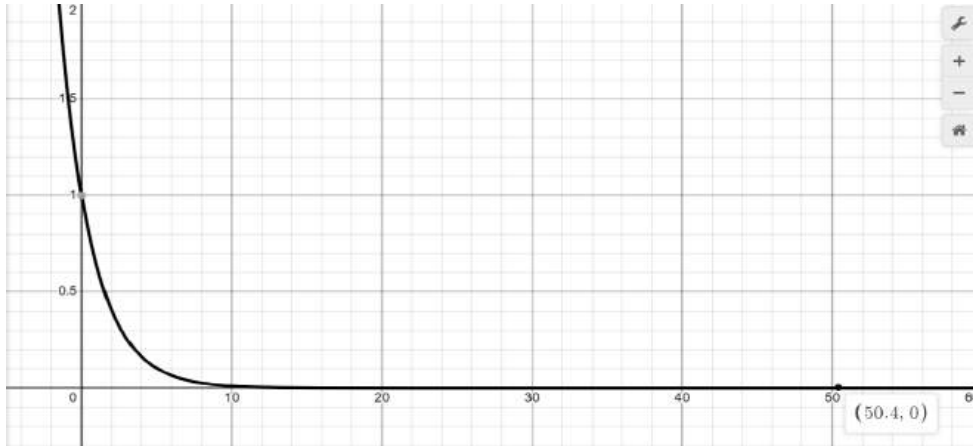


Fig 9. Graph to determine the percentage of muons given the time (non-relativistic)

Time is known to dilate near the speed of light. To determine the time dilation, the time from a stationary standpoint should be put in the context of a moving observer. In the equation $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$, known as the Lorentz factor, the gamma factor is produced [7]. This gamma factor is multiplied by the change in time to determine the time dilation. Following that, utilizing the lifetime equation of $I = e^{-t/\tau}$ and multiplying the gamma factor by the time will provide a new equation of $I = e^{-\frac{t\sqrt{1-\frac{v^2}{c^2}}}{\tau}}$ [7].

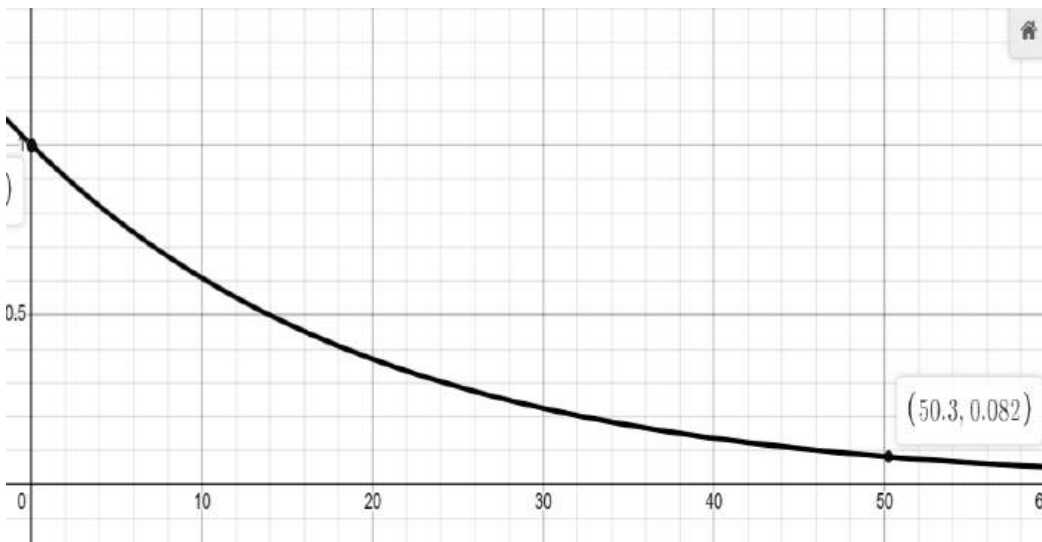


Fig 10. Graph to determine the percentage of muons given the time (relativistic)

This graph shows that under a relativistic framework, at a similar time traveled, the survival rate of muons is about 8.2%. Then using the gamma factor of approximately 0.994c as the relative velocity, the estimated time to travel this distance is 5.0582 μs based on the time dilation equation. Accounting for relativistic time dilation leads to more accurate estimations and preservation of unstable particles like muons.

Discussion

Muons can travel through different states of matter. A key benefit of the apparatus used is that it can be placed in different locations to estimate their presence. This allows us to envision unique applications, such as mapping the ocean floor by placing multiple apparatuses on the ocean bed. Once the apparatuses are positioned on the ocean floor, the time is recorded for the muons to the ocean floor. The difference in muon travel times between the surface and the ocean floor can be recorded, and with the muon speed already known, the distance between the ocean floor and the surface is determined. Repeated trials with portable and compact apparatuses can be done to efficiently map the constantly shifting ocean floors. Another benefit could be to capture the energy of muon decay. As the muon decays, a light flashes within the scintillator opening an opportunity to capture energy especially as thousands of muons travel through a medium in seconds. As we gain more understanding of the properties of this sub-atomic particle, exploration of new frontiers awaits.

Immunization in the US Army during World War II By Bella Gupta

Introduction

When the Second World War erupted, Americans watched as Europe plunged yet again into brutal warfare and bitter devastation. Naturally, the US Army immediately began talks and preparations for the possibility of American mobilization. The US recognized that in the case of deployment, disease would be a treacherous threat to soldiers. They knew that in wars past, disease ravaged armies without a shot ever being fired. Pathogens were known as “the other enemy” and had probably killed more soldiers than combat in the last 500 years of warfare. As recognized by the *Glasgow Courier* in 1943, “Disease was more deadly than bullets or winter.” Strict measures were needed to protect enlisted men, and new medical technology offered to change the entire course of the war. It was decided that vaccines could offer troops protection and curb the spread of disease in barracks and hospitals.

In the 1950s, the US Medical Department published nine volumes pertaining to preventative medicine during World War II. They serve as an unabridged guide to the role of medicine in the war. However since then, there have been few mentions or research done on the topic, and the significance is often ignored and overlooked. In this paper, we will discuss the development of vaccination campaigns across theatres, explore the diseases that proved most threatening to troops, ranging from typhoid to tetanus, and consider the significance of vaccines in strengthening the US Army and preventing the deaths of American soldiers during the war.

The Army Immunization Program

The army immunization program was one of the most necessary steps in vaccinating soldiers because it acted as an organized entity in charge of army immunization, from manufacturing to distribution and administration. The program began prior to the US entrance into World War II. In 1940, Brigadier-General James S. Simmons brought together a group of representatives of the Medical Services departments from the Army, Navy, and the Public Health Service to establish guidelines that would be “desirable in the event of war” (Preventative Medicine in World War II Vol. III by the US Medical Dept.). This precautionary meeting is understandable, considering the turmoil Europe became enthralled in following Hitler’s invasion of Poland.

The representatives under Simmons decided at this very meeting that the Armed forces would undergo tetanus immunization as quickly as possible, the first mandated vaccine of the war. The program soon became the bridge between the Army and Medical Department for the production, distribution, and decisions made pertaining to the vaccinations of troops, overseen by the Epidemiology Division of the Office of The Surgeon General. The government, the army, scientists, and civilian groups all cooperated for the collaborative effort to prevent the unnecessary death of soldiers throughout the war. The primary responsibilities of the program included:

Establishing policies for the immunization procedures; composing administrative instructions on vaccination campaigns; Improving vaccines and procedures circumstantially; Fixing issues that arise pertaining to adverse reactions from a vaccine; Distributing vaccines in accordance to storage guidelines; And conducting the correspondence between the multiple governmental, military, and civilian groups on the discussions for the immunization program.

The early tetanus immunization was one of multiple campaigns throughout this war, with some being continuations of routine vaccinations pre-war and others being newly adopted to confront diseases encountered in the Pacific. To produce such a large amount of vaccines, the government adopted “No loss, no gain” (Hoyt 2015) contracts, where institutions would be fully funded for vaccine development. The threat of war made scientists, universities, and manufacturers alike willing to work for little pay, as vaccines became essential to the war effort. Together, they were able to create, adapt, manufacture, and administer huge vaccination campaigns against the most dangerous diseases, protecting soldiers from the invisible threat facing them on the battlefield.

Typhoid-Paratyphoid

Typhoid fever is a bacterial infection that spreads through contaminated food and water, threatening troops in close, possibly unsanitary conditions. In the American Civil War, the fever was a leading cause of death for soldiers, but during World War I, immunization made typhoid rare and curbed the death rate remarkably. Administering the typhoid vaccine to all troops had been routine since 1911 and didn't need introduction as a new immunizing agent. Through the interwar years, advancements through research continued and brought about improvements to the vaccine, and production increased to accommodate the growing army. Precautionary measures arose in the early 1940s as war was in sight, and it was decided to re-adopt a typhoid-paratyphoid vaccine by the Army Immunization program to broaden protections against the disease. In September of 1940, the typhoid-paratyphoid vaccines were manufactured and were ready for distribution when needed. The vaccine itself was adapted to contain the English strain HA-6 in preparation for the possibility of troops being stationed in England. Other adaptations were made to best protect against the disease, and this particular vaccine strain was continuously manufactured throughout the war. However, it did not mean there were no incidences of typhoid within troops.

The effectiveness of this vaccine is displayed in the comparison between this war, with complete immunization, and the Spanish-American war, with no immunization practices. For example, in 1942, there were three cases of typhoid fever from American troops in India. This caused the shift to using an Indian-produced typhoid vaccine in this particular region for the remainder of the war. The US Medical Department declared that the routine use of typhoid vaccine was highly effective during the war due to the fact that there were only “one three-hundredths” of cases in the fully vaccinated army compared with the military during the

Spanish-American War. It is believed that the chances of being infected with typhoid were about the same from World War I to World War II, showing that typhoid vaccines were greatly beneficial against the spread of disease within troops. The China-Burma-India theatre had the highest rate of infection, which shows that the vaccine was probably most valuable in these regions, with most cases being men who had not yet completed a “basic series of typhoid vaccine.” From 1942 to 1945, 42 cases of typhoid fever were detected with certainty across all theatres, demonstrating how truly successful this vaccine campaign was.

Cholera

Cholera, like typhoid, spreads through contaminated water and food supplies. During the First World War, cholera’s spread caused intense fatigue, diarrhea, and vomiting in soldiers, preventing many from fighting for several days to weeks. However, in World War II, all troops stationed in Asia received the cholera vaccine, a relatively new procedure within the army but already used widely in South and Southeast Asia. A 1942 cholera outbreak of civilians in Madras, India, demonstrated that the rate of infection among the unvaccinated was “10.6 times” higher than in the vaccinated or “protected.” In China, there were about 13 cases of cholera and 2 deaths, attributed to contaminated food and water. However, this rate could have been much higher without the mass vaccination campaign against it. In 1945, there was an outbreak of cholera in Calcutta, and 50 British and Indian troops working alongside American soldiers became infected, with no cases occurring within vaccinated American troops (Preventative Medicine in World War II Vol. III by the US Medical Dept.). In India, the British army tended to vaccinate “only after an epidemic started,” not before, like the American troops, demonstrating that administering the vaccine beforehand most likely prevented American troops from falling ill with cholera in the Calcutta outbreak.

TABLE 10. DEATH RATES PER 1,000 PER ANNUM, DISEASE AND NONBATTLE INJURY, TOTAL UNITED STATES ARMY, 1900-45

Year	Disease	Nonbattle injury	Year	Disease	Nonbattle injury
1900	15.79	6.95	1923	2.01	1.90
1901	9.58	4.36	1924	1.94	1.89
1902	12.78	2.71	1925	1.77	1.99
1903	7.02	2.28	1926	2.27	1.67
1904	4.05	3.73	1927	2.35	1.65
1905	3.73	2.86	1928	2.30	1.80
1906	3.77	2.76	1929	2.29	2.05
1907	3.57	2.14	1930	2.08	1.81
1908	3.63	2.59	1931	2.47	2.13
1909	3.28	1.81	1932	2.11	2.15
1910	2.50	1.86	1933	2.11	2.10
1911	2.70	2.05	1934	2.04	1.99
1912	2.47	2.30	1935	2.09	1.75
1913	2.60	2.55	1936	2.00	2.03
1914	2.35	2.05	1937	1.87	1.61
1915	2.53	1.92	1938	1.54	1.71
1916	2.71	2.48	1939	1.55	1.50
1917	4.91	1.22	1940	1.04	1.76
1918	18.81	1.39	1941	.60	1.58
1919	7.61	1.34	1942	.68	2.08
1920	4.67	2.24	1943	.58	2.26
1921	2.24	2.02	1944	.55	2.45
1922	2.28	2.28	1945	.62	2.47

Source: Annual reports of The Surgeon General and "Army Battle Casualties and Nonbattle Deaths in World War II," Final Report, 7 December 41-31 December 46, AGO.

Figure 1. A table showing death rates from disease per 1000 troops per year from 1900-1945 Long, Arthur P. "Army Immunization Program ." *Preventative Medicine In World War II, III, Office Of The Surgeon General, Washington D.C., 1955, pp. 289-359.*

The Plague

Vaccines against the plague were administered on a case-by-case basis based on if or where a danger was detected. This meant that vaccines were stored with the troop's supplies and given if a threat was found. No cases of the plague were reported, nor was there any record of exposure, and in many cases, sanitation and environmental hygiene methods were as crucial as vaccine administration.

Typhus

Typhus, like typhoid, is a bacterial infection spread through contamination. During the first year of the war there was great and considerable advancement and development were made on the typhus vaccine to improve certain components for efficacy. Vaccination against typhus was received by all troops at risk of exposure, depending on the theatre. In 1943, a North African outbreak of typhus infected twelve American soldiers who were all vaccinated. However, every case was reported to be mild. This can be compared to the unvaccinated British troops, which had “three times as many cases, one-third of which ended fatally.” *Figure 1.* shows how beginning in 1941, the percentage of deaths from disease declined almost fifty percent, demonstrating how precautionary measures and vaccine campaigns prevented the death of thousands.

The next year, the National Research Council conference on Typhus issued a statement that “The evidence is all in favour...that while vaccination against epidemic louse-borne typhus fever does not always prevent infection, it causes the disease to be mild and appears thus far to prevent death from typhus” (US Medical dept. 1955). In 1942, a typhus outbreak broke out within unvaccinated Nazi forces in Poland. These troops were supposed to be vaccinated but “a mistake” (The Daily Monitor 1942) was made with production, and the Nazis were left unprotected. This demonstrates the importance of the organized procedures and protocols the US Army strictly adhered to regarding vaccine manufacturing and administration.



Women in the laboratory hard at work using egg-based production to develop an influenza vaccine
“*Influenza Vaccine Production at Commonwealth Serum Laboratories, 1950s.*” Omeka RSS,
cdcmuseum.org/exhibits/show/influenza/item/911. Accessed 25 Aug. 2024.

Influenza

The influenza virus is an airborne respiratory infection that spreads through coughing, sneezing, and proximity, making it a high risk for armed forces across any theatre. When the Spanish Flu pandemic began in 1918, American involvement in World War I was at its peak, which caused mass outbreaks in troops across Europe. Trenches were a breeding ground for disease to spread, and close, unsanitary conditions caused around “20% to 40% of US Army personnel” (Byerly 2010) to become infected with influenza. It killed more American soldiers than weapons ever did during the war. Countless men perished needlessly because of the rampant spread of the flu. And the men who didn’t expire from the disease were stricken with intense weakness and violent malaise. However, immunized soldiers played a great role in stopping another influenza tragedy during the Second World War. In 1940, doctors Jonas Salk and Thomas Francis created the first vaccine against influenza because of funding and support from the US Army, and they were administered to American troops. In 1943, the influenza vaccine procedure trials were given to certain commands, around twelve thousand personnel. However, the inoculations were reserved for outbreaks due to short supplies and a backlog in production, causing the mass vaccine mandate for every troop to not occur until late 1945. The timing of these 1943 trials was incredibly fortunate, considering they took place just two weeks before the 1943-1944 epidemic of Influenza type A. These trials were found to be marvellously successful, with one vaccinated man contracting the disease to every three and a half unvaccinated men contracting the disease. Following this particular administration was the 1945 outbreak, and while the widespread vaccination campaign did not have a control group like the previous epidemic outbreak, the low number of cases within troops can be compared to that of other service organizations. This comparison “indicated strongly” that the vaccines played a significant role in preventing deaths among army personnel during the 1945 Influenza B epidemic, avoiding the devastation the flu had brought the US Army just twenty-seven years earlier.



Lieutenant Marcella B. Brychta receiving the first flu vaccine in the India-Burma theatre during the war
“Discovering the Virus.” Omeka RSS,
cdcmuseum.org/exhibits/show/influenza/scientific-breakthroughs-giss/discovering-virus. Accessed 25 Aug.
2024.

Yellow fever

Yellow fever was of great concern in the Pacific theatre because of mosquito vectors. This fever was infamous for having ravaged the French army and forcing Napoleon’s retreat during his attempted re-conquest of Haiti in 1802. Before the American entrance into the war in 1940, Lt. Col. James S. Simmons worked with the Advisory Committee on Tropical Diseases to mandate yellow fever vaccinations to all military personnel in tropical regions before the Pearl Harbour attack. As an emergency measure, the International Health Division of the Rockefeller Foundation was requested to manufacture a large number of vaccines quickly. However, in 1942, an outbreak of jaundice proved to be a side effect from this vaccine, which caused the organization to pause administration and resume with a new vaccine provided by the Public Health Service that did not cause this symptom in recipients. This shows how mass immunization administrations during the war helped improve and develop new and safer vaccines. No troops contracted yellow fever in any of the tropical regions, in which immunization is said to have played a great role in. Evidently, through blood specimens, the medical department detected the presence of antibodies over four years after the date of vaccination, protecting soldiers actively in every region despite outbreaks near where troops were stationed.

Tetanus

Tetanus was a common risk in soldiers due to frequent battlefield and combat injuries. The Tetanus vaccine proved to be the most successful during the war period. The vaccine was first used within French troops during the ‘30s but was adopted by the US Army following its entrance into the war. The administration in France, which began in 1936, produced reports showing marvellous resistance to tetanus, with zero cases in French troops in 1940. Tetanus is an incredibly severe complication of wounds and surgery, especially on a battlefield. Because injury was a dire risk for almost all soldiers heading off to war, the Surgeon General advised immunization of all military personnel beginning in 1940. However, the War Department remained skeptical. Departments in the Philippines started to administer with great success anyway, and it was not until the War Department saw this that they approved of a widespread adoption across all theatres of war. “Undoubtedly... many lives were saved”, says the US Army Medical Department in the Pacific due to early usage of the tetanus toxoid vaccine. From 1941 to 1946, twelve soldiers contracted the disease, but for many of them, it was likely a result of not receiving a timely booster shot following their injury. Contrasting this to World War I, there were more than seventy cases between 1917 and 1918, showing a significant decline in outbreaks when the tetanus vaccine was used. During the D-Day invasion and aftermath in Normandy, the unvaccinated German ground soldiers were afflicted with eighty cases, while the vaccinated

Luftwaffe suffered no cases despite many injuries. This demonstrates how effective the vaccine was in protecting soldiers, proving to be a much better precaution than simply administering the antitoxin following injury. The antitoxin had many side effects, and by replacing it with the vaccine, many deaths were possibly avoided.

Conclusions

Vaccines were instrumental in keeping troops healthy and able to continue fighting thus changing the shape of the war in many ways. Soldiers not only faced threats from the battlefield but also from dangerous diseases, but the US military had significantly lower cases of almost every illness because of strict vaccine mandates that British and German forces lacked. The implementation of vaccines in the US Army not only prevented soldiers from infection but also began the routine practice of immunization for the civilian population against many diseases post-war. The tetanus vaccine, for example, became a regular inoculation for children in the US during the war, and since 1947, cases have declined by 95% (CDC 2024). This can be attributed to the rapid development and improvement of vaccines for the war effort and demonstrates how World War II prompted the funding for necessary scientific discoveries. Another example of how the war expanded civilian vaccination was with influenza. The first flu vaccine was created and used on troops through government funding before being approved for civilians in 1945. Over 100 million flu shots are given to Americans yearly, preventing hundreds of thousands of cases and providing preventative protection to the most vulnerable citizens.

One of the great ironies of this war is that while medical improvements like vaccines prevented the deaths of thousands of men and women, weapons of mass destruction were being brought to life. Vaccines were able to protect soldiers against bacteria and infection, but nothing could be done about the bitter combat and warfare those men faced. While vaccines were saving the lives of men across all theatres of the war, J. Robert Oppenheimer was in the midst of building the first atomic bomb, which ultimately killed around 250,000 people.

Nevertheless, the extraordinary role immunizations played in protecting the US Army against disease cannot be overstated, exceptionally in current times when medical misinformation is vehement, especially following the COVID-19 pandemic. Many soldiers viewed receiving their shot as their “patriotic duty” (Gambacorta 2021), and we must not forget that vaccines have and will continue to save the lives of many. The deprivation of World War II is inexplicable, but vaccines proved to save hundreds of thousands of American soldiers and strengthen the Allies in their victory. For that, the significance of vaccinations during the war can never be forgotten.

Works Cited

With guidance from Ian Olivo Read, PhD, professor of Latin American studies at Soka University

- “1885- a Mad Dog and a Fighting Frenchman!” Evening Star, 23 May 1943. Chronicling America,
<https://www.loc.gov/resource/sn83045462/1943-05-23/ed-1/?sp=94&q=vaccine+war&r=-0.217,0.484,1.738,0.792,0>.
- Byerly, Carol R. “The U.S. Military and the Influenza Pandemic of 1918-1919.” Public Health Reports (Washington, D.C. : 1974), U.S. National Library of Medicine, Apr. 2010, www.ncbi.nlm.nih.gov/pmc/articles/PMC2862337/.
- “Discovering the Virus.” Omeka RSS, cdcmuseum.org/exhibits/show/influenza/scientific-breakthroughs-giss/discovering-virus. Accessed 25 Aug. 2024.
- “Discovering the Virus.” Omeka RSS, cdcmuseum.org/exhibits/show/influenza/scientific-breakthroughs-giss/discovering-virus. Accessed 25 Aug. 2024.
- “Discovering the Virus.” Omeka RSS, cdcmuseum.org/exhibits/show/influenza/scientific-breakthroughs-giss/discovering-virus. Accessed 25 Aug. 2024.
- “GIs in U.S. Get Wonder Vaccine.” Detroit Sunday Times, 28 Oct. 1945. Chronicling America, <https://www.loc.gov/resource/sn88063294/1945-10-28/ed-1/?sp=11&q=vaccine+war&r=0.208,-0.039,0.63,0.287,0>.
- “Influenza Vaccine Production at Commonwealth Serum Laboratories, 1950s.” Omeka RSS, cdcmuseum.org/exhibits/show/influenza/item/911. Accessed 25 Aug. 2024.
- Kendall Hoyt Assistant Professor of Medicine. “How World War II Spurred Vaccine Innovation.” The Conversation, 7 Sept. 2023, theconversation.com/how-world-war-ii-spurred-vaccine-innovation-39903.
- Lal, Gobind Behari. “Can They Bomb Out The Armored T.B. Germ.” Detroit Evening Times, 14 May 1944. Chronicling America, <https://www.loc.gov/resource/sn88063294/1944-05-14/ed-1/?sp=54&q=germs+soldier&st=pdf&r=-0.286,-0.079,1.571,1.571,0>.
- Long, Arthur P. “Army Immunization Program .” Preventative Medicine In World War II, III, Office Of The Surgeon General, Washington D.C., 1955, pp. 289–359.
- “Medicine in World War I.” Yale University Library, onlineexhibits.library.yale.edu/s/wwi-medicine/page/diseases-at-the-battlefield. Accessed 25 Aug. 2024.
- “New Influenza Vaccine Developed in War To Be Ready Soon.” Evening Star, 23 Dec. 1945. Chronicling America, <https://www.loc.gov/resource/sn83045462/1945-12-23/ed-1/?sp=9&q=vaccine+war&r=0.094,0.066,0.639,0.291,0>.
- Ratcliff, J.D. “Germs As Zeppelins And Bombs Throw Doctors Into Confusion... And They Are More Deadly Than Shrapnel And High Explosives.” The Washington Herald, 31 Oct. 1915. Chronicling America,

- https://chroniclingamerica.loc.gov/lccn/sn83045433/1915-10-31/ed-1/seq-25/#date1=&index=10&rows=20&searchType=advanced&language=&sequence=0&words=pathogenic+soldier+soldiers&proxdistance=10&date2=&ortext=pathogen&proxtext=soldiers&phrasertext=&andtext=&dateFilterType=range&page=1&tab=tab_advanced_search.
- Ratcliff, J.D. "LIFESAVER." *Evening Star*, 5 Nov. 1944. *Chronicling America*,
<https://chroniclingamerica.loc.gov/lccn/sn83045462/1944-11-05/ed-1/seq-63/#date1=1935&index=3&rows=20&searchType=advanced&language=&sequence=0&words=soldier+Soldiers+soldiers+vaccine+vaccines&proxdistance=10&date2=1945&ortext=soldiers&proxtext=vaccine&phrasertext=&andtext=&dateFilterType=yearRange&page=1>.
- Ratcliff, J.D. "What Kills Soldiers? Germs More Deadly Than Bullets." *The Glasgow Courier*, 4 Feb. 1943. *Chronicling America*,
<https://www.loc.gov/resource/sn85042379/1943-02-04/ed-1/?sp=6&q=germs+soldier&r=0.149,0.425,0.707,0.322,0>.
- "Tetanus Surveillance and Trends." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, www.cdc.gov/tetanus/php/surveillance/index.html. Accessed 25 Aug. 2024.
- Turner, Matthew D, and Jason Sapp. "Failure to Plan: The Disease That Cost an American Empire." OUP Academic, Oxford University Press, 16 May 2023, academic.oup.com/milmed/article/188/7-8/171/7165307.
- "Typhus Menace to Axis Army." *The Daily Monitor*, 26 Oct. 1942. *Chronicling America*,
<https://www.loc.gov/resource/sn96077289/1942-10-26/ed-1/?sp=3&q=vaccine+war&st=image&r=0.021,0.298,0.478,0.218,0>.
- "Typhus Vaccine From America Reaches Far East." *Imperial Valley Press*, 26 June 1944. *Chronicling America*,
<https://www.loc.gov/resource/sn92070146/1944-06-26/ed-1/?sp=4&q=vaccine+war&r=0.282,-0.018,0.971,0.442,0>.
- "U.S. Army Readopts Triple Typhoid Inoculations." *Evening Star*, 30 Sept. 1942. *Chronicling America*,
<https://chroniclingamerica.loc.gov/lccn/sn83045462/1942-09-30/ed-1/seq-31/#date1=1935&index=0&rows=20&searchType=advanced&language=&sequence=0&words=soldier+Soldiers+soldiers+vaccination+vaccine+Vaccine&proxdistance=10&date2=1945&ortext=soldiers&proxtext=vaccine&phrasertext=&andtext=&dateFilterType=yearRange&page=1>.
- "U.S. Military Immunization: The Fight against Infectious Diseases." National Museum of Health and Medicine: The Micrograph - A Closer Look at NMHM, medicmuseum.health.mil/micrograph/index.cfm/posts/2019/us_military_immunization. Accessed 25 Aug. 2024.
- "War Medicine v.1 1941 Jan-Nov.." *HathiTrust*, babel.hathitrust.org/cgi/pt?id=mdp.39015070232585&seq=101&q1=influenza. Accessed 25 Aug. 2024.
- "War Medicine v.1 1941 Jan-Nov.." *HathiTrust*, babel.hathitrust.org/cgi/pt?id=mdp.39015070232585&seq=7. Accessed 25 Aug. 2024.
- "Shots' for Dogs." *Dickenson County Herald*, 8 Apr. 1943. *Chronicling America*,
<https://chroniclingamerica.loc.gov/lccn/sn95079130/1943-04-08/ed-1/seq-4/#date1=1935&index=8&rows=20&searchType=advanced&language=&sequence=0&words=soldiers+vaccine&proxdistance=10&date2=1945&ortext=soldiers&proxtext=vaccine&phrasertext=&andtext=&dateFilterType=yearRange&page=1>.

Simulation of Amyloid Beta Peptide Vaccine to Treat Alzheimer's **By Yasmina Kodeih**

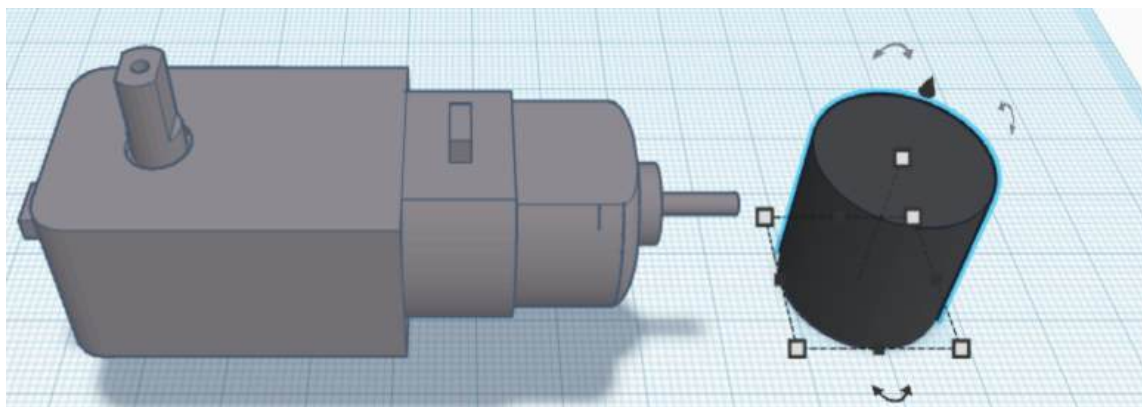
Introduction/Literature Review

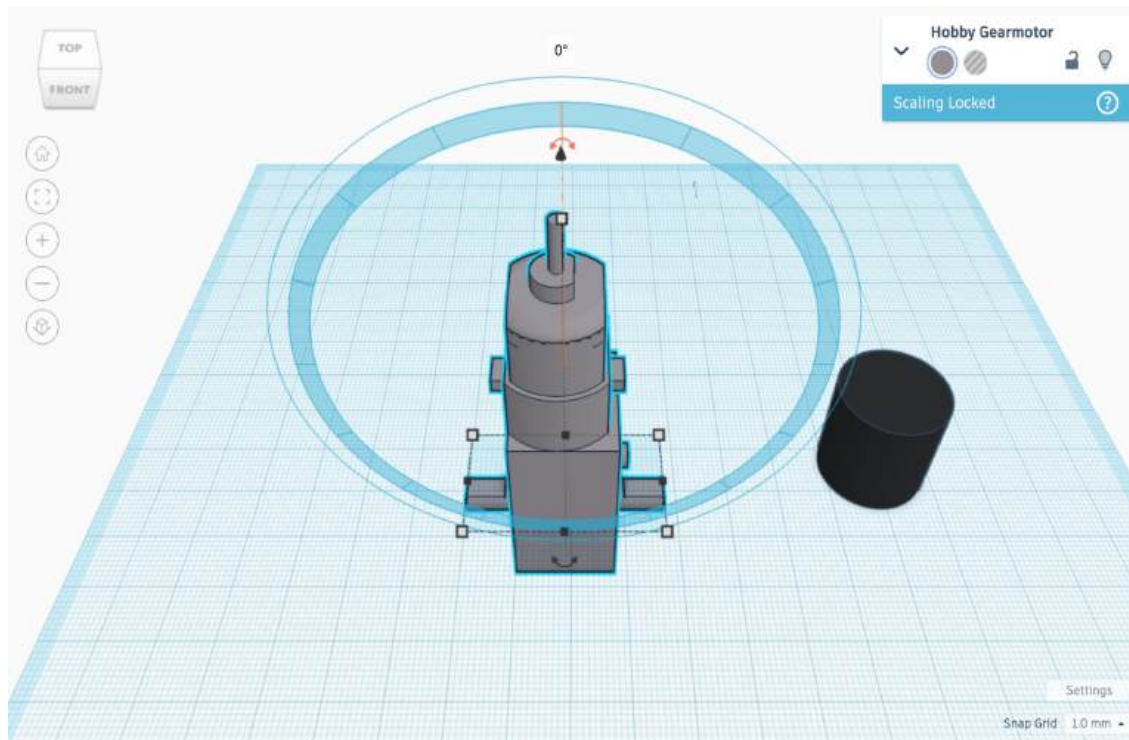
Alzheimer's disease is the sixth most and leading common cause of death in the United States. The seventh most frequent reason behind death for people 65 years of age or older. It was estimated that 5.8 million Americans who are 65 years of age or older will be diagnosed with Alzheimer's by 2020.¹ The cost of providing for them will reach \$360 billion by 2024. With the cost only expected to increase. Alzheimer's disease also might affect approximately 13 million Americans by 2050, and the disease will cost almost \$1 trillion. Through Medicaid and Medicare, the federal and state governments pay for two thirds of Alzheimer's disease-related expenses.² Alzheimer's disease is becoming more fatal than heart disease and cancer, which are both seeing decreases in death rates while Alzheimer's is seeing an increase. Research has shown that underreporting of Alzheimer's disease and other dementias on death certificates occurs often so the number of older people living with Alzheimer's disease may be much greater.¹ Alzheimer's disease is characterized as a brain condition that progressively reduces one's ability to function, think clearly, and remember moments or people. For most people with Alzheimer's the symptoms develop later in life.³ The idea that too much amyloid beta peptide formation in the brain is what causes Alzheimer's is the most scientifically supported explanation. 42 amino acids make up the amyloid beta peptide, which comes from the precursor protein, amyloid beta precursor protein (APP) which is expressed in numerous organs, particularly in the synapses of neurons. And a key player in the pathophysiology of Alzheimer's disease (AD). On chromosome 21, is the gene encoding of the amyloid beta precursor protein. The beta-secretase and gamma-secretase break apart the amyloid beta precursor protein to produce amyloid beta peptide, which has 40 or 42 amino acids and is designated as Ab 1-40 or Ab1-42. The peptide's N and C terminals are produced by beta secretase and gamma secretase on amyloid beta precursor proteins. Ab1-42 the primary pathogenic peptide is hydrophobic and has a propensity to form fibrils as well as oligomers. The hydrophobicity of the peptide is imparted to its C terminus by the amino acids that are hydrophobic. The amyloid plaques that are visible with the standard H and E staining of amyloid-specific staining are subsequently formed by these fibrils arranging in a beta-pleated sheet.⁴⁻⁵ Additionally, the buildup of amyloid beta protein causes protein clusters to develop. The main focus for the development of Alzheimer's treatment has been the A β Peptide.⁶⁻⁸ Many studies have been done manipulating the amyloid beta peptide in mice with varying success rates. But most show that amyloid beta peptide does in fact play a role.⁹⁻¹¹ And the creation of a vaccine to treat Alzheimer's has been discussed widely and shown to be a promising idea.¹²⁻¹⁵ However, the idea of reducing the amyloid beta in the brain via the use of a vaccine has not been looked into specifically using an endopeptidase (an enzyme than can break peptide bonds). This fact has created the project idea of creating a 3d model and simulation that shows a vaccine injecting an endopeptidase to break apart amyloid beta clusters in the brain. The reason a 3d model and simulation are most

appropriate at this time is because they are the perfect foundation for future creation and research into an amyloid beta vaccine. A 3d model demonstrates how a vaccine like this would look and be used. And a simulation proves whether or not the endopeptidase would be able to successfully break apart and reduce the formation of the amyloid beta clusters. By giving a starting point for a physical model of the vaccine and showcasing that a vaccine like this would in fact be successful allows for valuable future research/advancements on a amyloid beta vaccine. Also, even though my hypothesis is that the simulation will show that the endopeptidase is able to successfully break apart the amyloid beta, if it instead shows that an endopeptidase is not able to successfully break apart amyloid beta clusters in the brain. This project is still beneficial because it ensures that there is no time, effort, or resources wasted on creating an amyloid beta vaccine similar to the one created in this project as it will not work or will need changes in order to.

Methods

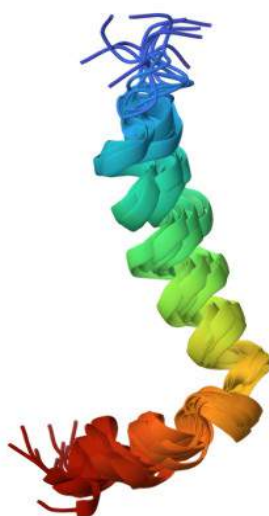
This project is a non-experimental creation project. As I am making a product not holding an experiment. The methods are mixed and this is because while my simulation will yield quantitative results my 3d model will not as I am not testing/counting anything. The reason I chose to make a product rather than hold an experiment is because at this stage of my academic career I am unable to hold any experiments on real Alzheimer's Patients. Also, I wanted to focus my project on advancing the research for vaccine creation. To create my 3d model I began by looking at common protein vaccine structures in order to ensure that my model would be accurate. I used the website Tinkercad (Tinkercad [Internet]; San Francisco: Autodesk, Inc.; c2024) to construct my model. I chose this website because I was familiar with it from secondary school. I began by creating an account with the website so as not to lose my progress then I watched the website tutorial video to refresh my memory. After I watched the video I created a new project then titled my project "Simulation Of AB Peptide Vaccine". I then started creating my model, I used the photos of vaccines, specifically protein vaccines, as a foundation and created a model I thought would get the job done. Here is it below.





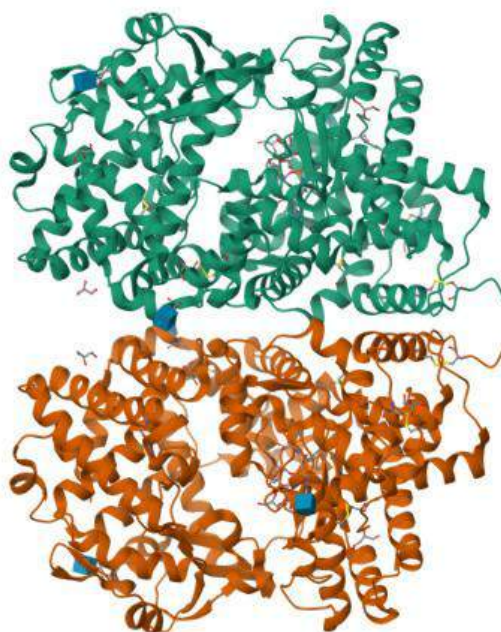
As you can see there are two grips on the side allowing the vaccine to be held between the pointer and middle finger and a button on the back that can be pressed with the thumb. There is a sharp point in the front of the vaccine which is intended to be the needle. The vaccine would be much thicker and larger than the average flu or COVID vaccine. This is because this vaccine would have to hold much more product than the average vaccine as the amount of endopeptidase needed would be larger than the amount of medicine in other vaccines. Also this vaccine would have to be injected in the brain (as stated the amyloid beta protein is in the brain) meaning that the structure of it would have to be much more complicated.

To create my simulation I began by researching the specific endopeptidase I wanted to use. The research in the field details neprilysin as a strong candidate so this is the endopeptidase that I chose.¹⁵⁻¹⁶ I decided to use the GROMACS application as it is free and I was slightly familiar with it unlike other simulation softwares. GROMACS is a software used for the simulation of nucleic acids, lipids, and proteins. To begin making the simulation I downloaded GROMACS on my personal computer I then watched a quick tutorial video to refresh my memory. After doing this I went into the "Protein Data Bank", which is a website that contains the molecular structure for all proteins. I searched "Amyloid Beta Protein " at the top right of the website. The correct photo for the protein popped up, it is below.



(Photo of Amyloid Beta from RCSB PDB(Protein Data Bank) was used in simulation)¹⁷

After I located the correct structure for the amyloid beta protein I clicked the “download” button on the upper right hand corner sending the download to a specific file I had created for the project. I then searched for the neprilysin structure and clicked the download button again, sending it to the same file. The photo is below.



(Photo of Substrate Free human Neprilysin from RCSB PDB(Protein Data Bank) was used in simulation)¹⁷

After I had both structures downloaded it was time to “clean them”. “Cleaning” the structures means preparing them to be sent into the simulation; this is necessary to ensure that there are no factors including the proteins that may affect the accuracy of the simulation results. To clean the protein you begin by removing all the water molecules and ligands this can be done with the tools within GROMACS (you pull the structure up and delete the water molecules and ligands the application indicates where these parts are). Then you ensure that the protein structure is accurate by pulling up photos of the structure and comparing. In my case the structures were accurate so I did not need to change them. It is also important to check for correct protein chain connectivity this means that the amino acids are connecting appropriately. To check this look at the residuals and bond angles they should be accurate for the structure. The correct values can be found with a google search. My values for both of my structures were accurate so I did not need to change them, if your values are not correct use a molecular modeling software to resolve the issues. Once my structure was “clean” it was time to build the simulation box. I used the Gromacs program “editconf:” which centers the proteins. I then prepared my submit script. My script included the module preparation and the Gaussian command. I used the script “esculentin.slm:”. I then used the “esculentin.slm:” script and “bsub” command in order to submit the job. I then monitored the job to ensure that everything was running smoothly. I ran the job 100 times to ensure that my results would be valid. After the job was done I needed to retrieve the output files, I ran the appropriate commands in the terminal (this is from the local directory where I wanted to download the files) and analyzed them. The main limitation in my method was my lack of simulation knowledge while I was familiar with GROMACS I had never run a simulation like this before. So there may have been more I could have done with the simulation that someone more knowledgeable would have known to do. A certain trick I was not aware of would have made my results more favorable or allowed me to run my simulation more efficiently which many have yielded different results.

Results/Analysis

My hypothesis was correct: the endopeptidase (neprilysin) was able to effectively break apart the amyloid beta in the brain. The results of my simulation showcased a 80% success rate out of 100 trials. This means that out of 100 trials Neprylsin was unable to break apart the amyloid beta clusters for 20 trials. This shows a relatively high success rate but indicates that it is not 100% successful. The simulation also indicated that in the successful trials it began to break apart the amyloid beta protein after about the equivalent of 4 or 5 days. These results indicate that a amyloid beta vaccine would work in reducing the amount of AB peptide clusters in the brain but do not indicate whether or not this would reduce or cure an individuals Alzheimer’s it is simply a treatment following the theory that the over formation of the amyloid beta protein in the brain is what causes Alzheimer’s. It also does prove or disprove the overformation theory, it just gives a treatment for it.

Conclusion

The results of this project indicate that the endopeptidase, neprilysin specifically, is able to efficiently break apart the amyloid beta clusters in the brain. Meaning that if the theory that the reformation of the amyloid beta clusters is what causes Alzheimer's is true this vaccine, if advanced upon, could be used as a treatment for Alzheimer's. But again it does indicate whether or not the theory is true. The limitation present in this project is that the ability for neprilysin to break apart the amyloid beta clusters in the brain was run on a simulation. The human brain is complicated and not as predictable as a simulation. So it is possible that if neprilysin is used to break apart the amyloid beta clusters in a real human brain the success rate will not be exactly 80%. Also the 3d model created may not be able to be replicated physically very easily as it is just a model. The implications of this project are that more research will be done into the creation of an amyloid beta vaccine to treat Alzheimer's as this project shows it has promise. This project also adds to the research on the over formation of AB peptide causing Alzheimer's theory. Which in turn can help us determine where this theory is true or not. For those attempting to replicate or advance this research I recommend beginning to create a physical model of the amyloid beta vaccine that is based on the 3d model shown in this project. I also recommend running the simulation with a different endopeptidase. Even though neprilysin was the endopeptidase that was the most discussed in my field as a potential amyloid beta breaker there are many others. And these other endopeptidases may yield better results; examples of these other endopeptidases are Proteolysis-Targeting Chimera Strategy or an Insulin-Degrading Enzyme (IDE).

Works Cited

1. What is Alzheimer's disease? Centers for Disease Control and Prevention. October 26, 2020. Accessed April 30, 2024. <https://www.cdc.gov/aging/aginginfo/alzheimers.htm>.
2. Public Health Approach to Alzheimer's. Alzheimer's Disease and Dementia. Accessed April 30, 2024. <https://www.alz.org/professionals/public-health/public-health-approach>.
3. Alzheimer's disease fact sheet | National Institute on Aging. National Institute on Aging (NIA). Accessed April 30, 2024. <https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet>.
4. Rukmangadachar LA. Amyloid beta peptide. StatPearls [National Library Of Medicine]. August 28, 2023. Accessed April 30, 2024. <https://www.ncbi.nlm.nih.gov/books/NBK459119/>.
5. Chen G, Xu T, Yan Y, et al. Amyloid beta: Structure, biology and structure-based therapeutic development. Nature News. July 17, 2017. Accessed April 30, 2024. [https://www.nature.com/articles/aps201728#:~:text=Amyloid%20beta%20peptide%20\(A%CE%B2\)%20is,%CE%B2%2D%20and%20%CE%B3%2Dsecretases](https://www.nature.com/articles/aps201728#:~:text=Amyloid%20beta%20peptide%20(A%CE%B2)%20is,%CE%B2%2D%20and%20%CE%B3%2Dsecretases).
6. Zhang Y, Chen H, Li R, Sterling K, Song W. Amyloid β -based therapy for alzheimer's disease: Challenges, successes and future. Nature News. June 30, 2023. Accessed April 30, 2024. [https://www.nature.com/articles/s41392-023-01484-7#:~:text=Amyloid%20%CE%B2%20protein%20\(A%CE%B2\)%20is,the%20development%20of%20AD%20therapy](https://www.nature.com/articles/s41392-023-01484-7#:~:text=Amyloid%20%CE%B2%20protein%20(A%CE%B2)%20is,the%20development%20of%20AD%20therapy).
7. Nordberg A. Amyloid plaque imaging in vivo: current achievement and future prospects. European Journal of Nuclear Medicine & Molecular Imaging. 2008;35:46-50. doi:10.1007/s00259-007-0700-2
8. Author links open overlay panelFuxiang Bao a 1, a, 1, et al. Different β -amyloid oligomer assemblies in alzheimer brains correlate with age of disease onset and impaired cholinergic activity. Neurobiology of Aging. June 17, 2011. Accessed April 30, 2024. <https://www.sciencedirect.com/science/article/abs/pii/S0197458011001709?via%3Dihub>.
9. Qian X, Liu X, Chen G, Chen S, Tang H. Injection of amyloid- β to lateral ventricle induces gut microbiota dysbiosis in association with inhibition of cholinergic anti-inflammatory pathways in alzheimer's Disease - Journal of Neuroinflammation. BioMed Central. September 28, 2022. Accessed April 30, 2024. <https://jneuroinflammation.biomedcentral.com/articles/10.1186/s12974-022-02599-4>.
10. Faucher P, Mons N, Micheau J, Louis C, Beracochea DJ. Hippocampal injections of oligomeric amyloid β -peptide (1–42) induce selective working memory deficits and long-lasting alterations of Erk Signaling pathway. Frontiers. December 14, 2015. Accessed April 30, 2024. <https://www.frontiersin.org/articles/10.3389/fnagi.2015.00245/full>.
11. Vicidomini C, Borbone N, Roviello V, Roviello GN, Oliviero G. Summary of the current status of DNA vaccination for alzheimer disease. Vaccines. November 10, 2023. Accessed April 30, 2024. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10674988/>.

12. Ananthaswamy A, Klein A. Could a vaccine beat Alzheimer's? *New Scientist*. 2016;230(3076):10. doi:10.1016/S0262-4079(16)30981-2
13. Adi - The Future of vaccines for alzheimer's and dementia. Alzheimer's Disease International (ADI). June 29, 2023. Accessed April 30, 2024. <https://www.alzint.org/resource/the-future-of-vaccines-for-alzheimers-and-dementia/>.
14. Researchers return to alzheimer's vaccines, buoyed by recent drug success | Reuters. Accessed April 30, 2024. <https://www.reuters.com/business/healthcare-pharmaceuticals/researchers-return-alzheimers-vaccines-buoyed-by-recent-drug-success-2023-11-20/>.
15. 1. El-Amouri SS, Zhu H, Yu J, Marr R, Verma IM, Kindy MS. NEPRILYSIN: An enzyme candidate to slow the progression of alzheimer's disease. *The American journal of pathology*. May 2008. Accessed April 30, 2024. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2329843/>.
16. Brum WS, Docherty KF, Ashton NJ, et al. Effect of NEPRILYSIN inhibition on alzheimer disease plasma biomarkers: A secondary analysis of a randomized clinical trial. *Lunds universitet*. February 7, 2024. Accessed April 30, 2024. <https://portal.research.lu.se/sv/publications/effect-of-neprilysin-inhibition-on-alzheimer-disease-plasma-bioma>.
17. Bank RPD. 1IYT: Solution structure of the alzheimer's disease amyloid beta-peptide (1-42). RCSB PDB. Accessed April 30, 2024. <https://www.rcsb.org/structure/1iyt>.

The Evolution, Impacts, and Future Prospects of Genetically Modified Organisms

By Xiaoyu Zou

Abstract

Genetically Modified Organisms (GMOs) now cover an estimated 190 million hectares globally, underscoring their vast impact on modern agriculture, biotechnology, and environmental sustainability. Originating from the pioneering advances in recombinant DNA technology during the early 1970s, GMOs have significantly transformed agricultural practices by enhancing crop yield, resilience, and nutritional value. This review traces the progression from early genetic engineering breakthroughs to the advanced CRISPR-Cas9 system, highlighting the precision and efficiency of modern GMO technologies. While the benefits include improved food security and reduced environmental degradation, GMOs also spark debates over safety, ethical implications, and socio-economic disparities. Case studies such as Golden Rice and Bt cotton illustrate these complexities, emphasizing the nuanced integration of GMOs into global agriculture. Looking ahead, GMOs hold promise in addressing critical global challenges such as climate change and food distribution inequities, necessitating a balanced approach to capitalize on their benefits while mitigating risks. This review aims to contribute to ongoing debates and deepen the understanding of GMOs' pivotal role in contemporary society, advocating for continued research and the development of fair policy frameworks.

Introduction

Genetically Modified Organisms (GMOs) have become a cornerstone of modern agriculture and biotechnology, revolutionizing the way we cultivate crops, raise livestock, and even approach healthcare (Oliver, 2014). This literature review delves into the multifaceted world of GMOs, providing a detailed analysis of their development, impacts, and the potential that lies ahead. Through genetic engineering, the very blueprint of organisms can be altered to exhibit traits that enhance their growth, nutritional value, and resistance to pests and diseases (Phillips, 2008). This manipulation not only has significant implications for food production but also for environmental sustainability, economic growth, and the global food supply chain.

GMOs are defined as organisms whose genetic material has been altered in a way that does not occur naturally by mating or natural recombination. This process, known as genetic engineering, involves the insertion or deletion of genes to achieve desired traits such as increased yield, drought resistance, or enhanced nutritional content. The advent of GMOs marked a significant departure from traditional breeding techniques, offering more precision and efficiency in developing organisms with specific beneficial characteristics (Phillips, 2008).

The global trend in GMO cultivation and usage has seen a remarkable upsurge over the past few decades. As of the latest statistics, millions of hectares worldwide are devoted to genetically modified crops, with a significant proportion in countries like the United States, Brazil, Argentina, Canada, and India (Kamle et al., 2017). These crops primarily include soybean, corn, cotton, and canola, which have been genetically modified for herbicide tolerance, insect resistance, or both. The

proliferation of GMO cultivation underscores the growing reliance of the agricultural sector on biotechnological innovations to meet the food demands of a burgeoning global population.

The importance of GMOs in modern agriculture and biotechnology cannot be overstated. They offer a promising solution to many of the challenges facing contemporary food production systems, including pest infestations, climate change, and soil degradation (Oliver, 2014). By improving crop resilience and reducing the dependence on chemical pesticides and fertilizers, GMOs contribute to environmental sustainability. Furthermore, they hold the potential to address nutritional deficiencies in developing countries through biofortified crops, exemplifying their role in enhancing food security and public health (Sandhu, et al. 2023). As we advance, the future prospects of GMOs hinge on navigating ethical, environmental, and regulatory challenges, ensuring that biotechnological innovations continue to serve the global community's needs sustainably and responsibly. This literature review aims to explore these aspects in depth, offering insights into the ongoing debates and research surrounding GMOs.

History and Development of GMOs

The history and development of Genetically Modified Organisms (GMOs) is a fascinating journey that spans several decades, marked by groundbreaking research and significant technological advancements. This section of the paper delves into the early research and breakthroughs in genetic engineering that paved the way for the creation of GMOs, traces the evolution of GMO technology from its inception to its current state, and highlights the major milestones in the development of GMO crops, including the first instances of regulatory approvals.

Early Research and Breakthroughs in Genetic Engineering Leading to GMOs

The foundation of genetically modified organism (GMO) technology was laid in the early 1970s with the advent of recombinant DNA (rDNA) technology, pioneered by scientists Herbert Boyer and Stanley Cohen (Boyer 1972). This groundbreaking method involves extracting a desired gene from a donor organism, using restriction enzymes to cut and remove it, and then inserting it into a vector—typically a plasmid—that carries the gene into the host organism's cells. Through processes like transformation, where the recombinant DNA is introduced into the host, and techniques such as gene guns or electroporation, the DNA is integrated into the host's genome (Boyer, 2024). This allows the host to express the new trait encoded by the inserted gene, such as pest resistance or environmental stress tolerance. The development of rDNA technology marked a significant leap, setting the stage for further innovations in genetic engineering by demonstrating the potential to manipulate genetic material to develop organisms with desirable traits efficiently.

Evolution of GMO Technology from Inception to Current State

The evolution of genetically modified organism (GMO) technology has reached a pivotal milestone with the advent of the CRISPR-Cas9 system. Derived from a bacterial defense mechanism, CRISPR-Cas9 has transformed genetic engineering with its ability to make precise

DNA edits efficiently and flexibly. This system utilizes a guide RNA to direct the Cas9 enzyme to a specific DNA sequence, where it creates a targeted cut. The cell then repairs this cut via natural mechanisms like non-homologous end joining (NHEJ) or homology-directed repair (HDR), the latter allowing for the insertion of new DNA sequences at the site of the cut (Rodríguez-Rodríguez et al., 2019). Compared to older gene-editing technologies such as zinc finger nucleases (ZFNs) and TALENs, which require intricate protein engineering, CRISPR-Cas9's reprogrammable guide RNA offers a simpler, more cost-effective, and precise method. This has significantly sped up genetic research and enhanced the development of GMOs, improving crop yields, pest resistance, and nutritional values, thereby setting the stage for future agricultural and biotechnological innovations. (Castro et al., 2021)

Major Milestones in GMO Development

The journey of GMO development is marked by several major milestones. One of the earliest successes was the creation of the Flavr Savr tomato in the early 1990s, the first GMO crop to receive regulatory approval for commercial production (Maurer, 2015). Engineered to have a longer shelf life, the Flavr Savr tomato represented a significant step forward in the application of genetic engineering in agriculture.

Subsequently, a range of GMO crops received regulatory approval, including herbicide-resistant soybeans and insect-resistant corn, which have become widely adopted in many countries around the globe. These developments were accompanied by the establishment of regulatory frameworks to assess the safety and environmental impact of GMOs, ensuring that genetically modified crops meet stringent standards before reaching the market (Oliver, 2014).

The development of GMOs has not been without controversy, with debates surrounding their safety, environmental impact, and ethical considerations. However, the scientific consensus supports the safety of GMOs for human consumption and their potential benefits for agriculture and food security.

In summary, the history of GMO development is a testament to the remarkable progress in genetic engineering technology. From the initial breakthroughs in recombinant DNA technology to the sophisticated genome editing tools of today, the journey of GMOs reflects a broader narrative of scientific advancement and innovation. As we look to the future, the ongoing evolution of GMO technology holds the promise of further contributions to sustainable agriculture, food security, and the development of new solutions to global challenges.

Effects of GMOs on Human Health

The impact of genetically modified organisms (GMOs) on human health remains a critical area of study within the fields of biotechnology, nutrition, and public health. The effects of GMOs can be examined from both advantageous and disadvantageous perspectives, particularly in how they influence nutritional standards and potentially elicit adverse health reactions.

Advantages: Enhancing Nutritional Standards Through GMOs

One of the most promising advantages of GMOs is their ability to enhance nutritional content in foods. Genetic modification has been employed to fortify crops with essential vitamins, minerals, and proteins that may be lacking in the diets of populations, especially in developing countries (Sushil 2019).

Additionally, there is ongoing research into developing edible vaccines through genetically modified crops, where specific genes responsible for vaccine antigens are inserted into the plant's genome. These plants could then theoretically be used as vaccines that are easily administered, reducing the need for conventional vaccine production and distribution methods, which can be costly and require more complex logistics (Anjana, 2018).

Further advancements in genetic modification have led to the production of GMO crops that lack common allergens or natural toxins. For instance, research has been directed toward developing peanut plants that lack key allergenic proteins, potentially reducing the risk of life-threatening allergic reactions among sensitive individuals (Richard E. 2024).

Disadvantages: Potential Adverse Health Reactions and Regulatory Challenges

Despite these benefits, there are concerns regarding the unintended consequences of consuming GMO foods. The inclusion or removal of certain genes might lead to unexpected adverse health reactions in some individuals. One documented example illustrating the potential for genetically modified organisms to produce unintended allergens involves the experimental GMO soybeans that were genetically engineered with a gene from Brazil nuts. The intention was to enhance the soybeans with the amino acid methionine for greater nutritional value (Inna, et al 2023). However, during development, it was discovered that the transgenic soybeans elicited an allergic reaction in individuals allergic to Brazil nuts. This discovery, made prior to commercial release, led to the cessation of the project to prevent potential health risks to consumers. This case highlights the importance of thorough allergenicity testing in GMO development to ensure safety and prevent unforeseen adverse health reactions (Inna, et al 2023).

A significant issue with GMOs in the consumer market is the lack of mandatory labeling in some countries, which makes it difficult for consumers to make informed choices about what they eat (Wunderlich, 2015). Without clear labeling, consumers may unknowingly consume products containing GMOs, which is a concern for those with specific dietary restrictions, allergies, or philosophical objections to genetically modified foods.

Case Study 1: Golden Rice

A prime example of this is "Golden Rice," which has been genetically engineered to produce beta-carotene, the precursor to vitamin A. Vitamin A deficiency is rampant in many parts of Asia and Africa, leading to serious health issues such as blindness and increased mortality rates. Golden Rice offers a potential dietary supplement that could help alleviate these health problems (Mongabay 2023).

Background and Development

Golden Rice represents a pivotal development in the field of genetically modified organisms (GMOs) aimed at addressing critical global health issues. It was developed in the late 1990s by Professors Ingo Potrykus of the Swiss Federal Institute of Technology and Peter Beyer of the University of Freiburg. The primary goal of Golden Rice was to combat vitamin A deficiency (VAD), a severe public health issue affecting millions of children and pregnant women in over half of all countries, especially in Africa and Southeast Asia. VAD is associated with an increased risk of blindness, susceptibility to disease, and higher mortality rates Golden Rice is (Marci, et al).

genetically modified to produce beta-carotene, a precursor of vitamin A, in the rice endosperm—the part of the rice that people eat. This was achieved by inserting genes from a daffodil (and later from maize, which proved more effective) that enable the biosynthesis of beta-carotene in the rice grain (Tang, 2009). The name "Golden Rice" is derived from the golden color of the rice, which results from the beta-carotene content.

Health Benefits

The health benefits of Golden Rice are potentially significant. By providing a dietary source of vitamin A, Golden Rice could help improve the nutritional status of populations in many developing countries where rice is a staple food, and vitamin A-rich foods are scarce or too expensive for the average household. The World Health Organization estimates that 250,000 to 500,000 vitamin A-deficient children become blind every year, half of whom die within a year of losing their sight. In these contexts, Golden Rice could act as a supplementary source of vitamin A and significantly mitigate the impacts of VAD (Tang, 2009).

Challenges and Controversies

Despite its potential benefits, the deployment of Golden Rice has been fraught with challenges and controversies:

Effectiveness and Practicality: Critics have questioned the effectiveness of Golden Rice as a solution to VAD. They argue that the amount of Golden Rice needed to be consumed to meet daily vitamin A requirements could be impractically high (Kettenburg, 2018). Additionally, the beta-carotene in Golden Rice degrades over time, and its bioavailability can be affected by the overall diet of the consumers, including their fat intake, which is necessary for the absorption of vitamin A.

Environmental and Health Safety: Concerns have been raised about the potential environmental impact of cultivating a genetically modified crop widely across different ecosystems. Critics worry about the potential for gene flow to wild rice varieties and the unintended consequences that might arise from this, such as affecting biodiversity. Health safety concerns also include the long-term effects of consuming GM foods, which some groups argue have not been adequately studied (Kettenburg, 2018).

Intellectual Property and Accessibility: Golden Rice has also been a focal point of debates over intellectual property rights. Initially, various patents covered the technologies used in its development, which could restrict access to those who need it most—smallholder farmers in developing countries. However, the inventors and involved institutions have made efforts to offer Golden Rice to resource-poor farmers free of charge, under humanitarian licenses, provided it is used for non-commercial purposes.

Public Perception and Acceptance: The acceptance of Golden Rice by local communities is crucial for its success. However, skepticism about GMOs among the public and within various non-governmental organizations (NGOs) has led to resistance and delays in adoption. Campaigns both for and against Golden Rice often reflect broader global debates over the use of genetically modified crops.

Current Status

Despite these challenges, efforts to introduce Golden Rice continue, with field trials and regulatory assessments being conducted in countries like Bangladesh and the Philippines. These trials aim to evaluate the effectiveness and safety of Golden Rice under local conditions, an essential step toward wider acceptance and use.

In summary, Golden Rice exemplifies the complexities and potential of genetic engineering to address significant health issues but also highlights the myriad technical, ethical, and social challenges that must be navigated to realize these benefits.

Effects of GMOs on Environment

The impact of genetically modified organisms (GMOs) on the environment is a subject of significant debate and analysis within the fields of ecology, agriculture, and environmental science. The introduction of GM crops has been associated with both potential environmental benefits and risks. Understanding these impacts is crucial for developing responsible practices and policies regarding the use of GMOs.

Disadvantages of GMOs on the Environment

One of the significant environmental concerns related to GMOs is their potential impact on biodiversity. GMOs, engineered to possess advantageous traits such as higher yields or resistance to pests, can outcompete indigenous plant life, leading to a reduction in biodiversity. This occurs when these modified plants monopolize resources like sunlight, water, and soil nutrients, displacing native species and reducing the variety of plants within the ecosystem (Landry 2015). Additionally, the ecological balance can be further disturbed through unintended effects on non-target species. For instance, proteins or toxins produced by GM crops designed to target specific pests might inadvertently affect other organisms in the ecosystem, such as beneficial insects or soil microbes, potentially disrupting existing food webs and ecological functions (Ladics, et al. 2015).

Another pressing issue is the risk of creating herbicide-resistant weeds and accelerated evolution of resistant pest species. As farmers adopt herbicide-resistant crops, there is a tendency to increase the usage of specific herbicides, leading to selective pressure that favors the survival of resistant weed species. Over time, these resistant populations can become dominant, complicating weed management and increasing agricultural reliance on herbicides (Gaines, et al.). Similarly, continual exposure to the biopesticides produced by GM crops can lead to the development of resistant pest populations, a phenomenon akin to antibiotic resistance in bacteria, necessitating the development of new agricultural strategies and technologies(Siddiqui, et al).

Advantages of GMOs on the Environment

Conversely, GMOs also offer environmental benefits, notably in reducing the need for chemical inputs such as insecticides. By engineering crops that are inherently resistant to pests, the use of chemical insecticides can be significantly decreased. This reduction is beneficial to the environment as it leads to lower levels of chemical runoff into rivers and soils, thereby preserving the health of aquatic life and maintaining soil quality (Oliver 2015). Additionally, GM crops can be designed to thrive in sub-optimal conditions, such as areas with poor soil quality or lower water availability. This capability potentially reduces the need to clear forests for agricultural purposes, aiding in the conservation of natural habitats and contributing to carbon sequestration efforts(Oliver, 2014).

Case Study: Bt Cotton and Its Environmental Impact

Bt cotton represents a significant innovation in genetically modified crops, engineered specifically to combat pervasive pests in cotton farming. By the toxin producing gene from the bacteria (Bt), known for its insecticidal properties, into cotton plants, Bt cotton directly addresses the economic and environmental challenges posed by pests like the bollworm (Abbas, 2018).

Background and Development

Developed in the 1990s, Bt cotton was engineered to produce the Bt toxin, an insecticidal protein that targets specific pests, notably the bollworm. This genetic modification aimed to reduce the reliance on chemical pesticides, thus decreasing production costs and minimizing environmental damage from pesticide overuse. The introduction of Bt cotton marked a pivotal shift in agricultural biotechnology, offering farmers an integrated pest management tool that could significantly enhance yield and reduce crop damage(Abbas, 2018).

The development of Bt cotton involved the insertion of a gene from *Bacillus thuringiensis* into the cotton genome. This gene codes for a protein that is toxic to certain insects but is considered safe for humans and other mammals. The genetic modification allows the cotton plant to produce this toxin unnaturally, effectively protecting itself from specific insect pests(Abbas, 2018). For example Cotton Bollworm which is also known as the corn earworm or tomato fruitworm in various regions, this pest is prevalent in many areas where cotton is grown and can cause extensive damage to the cotton bolls, the fruiting bodies of the cotton plants.

Environmental Benefits

The primary environmental benefit of Bt cotton has been the substantial reduction in the use of chemical insecticides. In regions where Bt cotton has been widely adopted, such as India and the United States, there has been a marked decrease in pesticide usage. For instance, a study noted that in India, pesticide use on cotton dropped by an estimated 50% from 2000 to 2010 following the introduction of Bt cotton (Peshin et al. 2021). This reduction not only lowers the ecological footprint of cotton farming but also diminishes the risk of pesticide runoff into water bodies and soil, which can harm aquatic life and degrade soil health. Furthermore, by decreasing the need for mechanical pest control, Bt cotton contributes to lower fuel consumption and less soil compaction. Specific data from a study in 2012 showed that Bt cotton fields in Arizona required fewer tractor passes for pest control, leading to a reduction in fuel use by about 2.5 liters per hectare and significantly less soil compaction, thereby promoting a more sustainable agricultural practice (Kathage, 2012).

Environmental Challenges

Despite its benefits, the environmental impact of Bt cotton includes several challenges. One significant concern is the potential for pest resistance. Continuous exposure to the Bt toxin can lead to the development of resistance among target pest populations, a phenomenon observed in some regions where Bt cotton has been extensively cultivated. This resistance necessitates the development of new strategies or the introduction of new Bt genes to maintain efficacy.

Additionally, there have been reports of secondary pests emerging as significant problems once the primary pests are controlled by Bt cotton. This shift in pest dynamics can sometimes lead to an increase in the use of other pesticides, offsetting some of the initial environmental benefits of Bt cotton.

Current Status

Today, Bt cotton is cultivated in numerous countries, and it continues to play a crucial role in integrated pest management strategies. However, the sustainability of Bt cotton is under scrutiny due to issues of pest resistance and secondary pest outbreaks. Ongoing research focuses on stacking multiple Bt genes in cotton plants to delay resistance and developing integrated pest management practices that include crop rotation and the use of non-Bt refuges to manage resistance.

In summary, Bt cotton illustrates both the potential and the limitations of genetically modified crops in environmental management. While it offers significant advantages in terms of reduced pesticide use and enhanced crop protection, it also presents challenges that require careful management and continuous scientific innovation to ensure long-term sustainability and ecological balance.

Economic Consequences of GM Crops

The economic implications of genetically modified organisms (GMOs) are a critical area of study in agricultural economics, policy-making, and global trade. The integration of GM crops into

agriculture has brought about significant economic shifts, offering both remarkable opportunities and complex challenges. These economic impacts influence a wide range of stakeholders, including farmers, consumers, biotechnology companies, and national economies. Understanding the economic consequences of GMOs is essential for developing balanced agricultural policies and practices that maximize benefits while addressing potential downsides.

Advantages

Genetically Modified (GM) crops offer several economic benefits that stem from their enhanced agricultural traits. One of the primary advantages is the increased yield that these crops can provide. By incorporating genes that enhance resistance to pests and diseases, GM crops suffer less damage and loss compared to conventional crops, thereby significantly increasing yield (Oliver, 2014). Moreover, GM crops are often engineered to withstand environmental stresses such as drought, salinity, and extreme temperatures, which not only ensures yield stability but also expands the areas where crops can be viably cultivated, thus increasing land use efficiency (Aziz, et al 2022).

Another significant economic benefit for farmers is the potential for reduced farming costs. GM crops designed to be herbicide-resistant allow for more straightforward weed management, which can lower the costs of herbicides and the labor required for their application. Additionally, crops engineered to have longer shelf lives, such as GM tomatoes with delayed ripening traits, can reduce losses due to spoilage, enhancing profitability for producers and distributors. This reduction in post-harvest losses is particularly crucial in regions with less developed logistics and storage infrastructure (Brookes 2022).

Disadvantage

While GM crops offer considerable economic benefits, they also raise significant issues related to the patent system governing their use. Most GM seeds are patented by biotech companies, which can restrict farmers' traditional practices of saving and replanting seeds from one year to the next (Sciences). Farmers must purchase new seeds each season, often at a higher cost, which can increase their dependence on seed companies and reduce their operational autonomy.

The patent system can also impact seed diversity. The dominance of patented GM seeds in the market can lead to a decrease in the cultivation of traditional and local crop varieties, which may have been developed over generations to adapt to specific local conditions. This reduction in agrobiodiversity can make food systems more vulnerable to pests, diseases, and changes in climate because fewer varieties mean that crops are more genetically uniform across large areas (Sciences).

In summary, while GM crops can lead to increased yields, enhanced resilience to environmental stresses, and reduced farming costs, they also present challenges through the patent system that can affect farmers' rights and seed diversity. These economic impacts require careful consideration and management to ensure that the benefits of GM technologies are balanced with sustainable agricultural practices and the preservation of genetic diversity.

Case Study: GM Soybeans in Argentina

Background and Development

Genetically modified (GM) soybeans were introduced to Argentina in the mid-1990s, primarily to enhance productivity and resistance to herbicides, especially glyphosate. Argentina rapidly adopted GM soybean varieties due to their promise of higher yields and easier weed control (Kunin et al 2024). This adoption transformed Argentina into one of the world's leading soybean producers, significantly affecting the nation's agricultural landscape and economy.

The development of GM soybeans involved the incorporation of genes that conferred resistance to glyphosate, allowing farmers to apply this herbicide without damaging the crop. This technology was initially developed by Monsanto and found rapid acceptance in Argentina, where soybeans play a crucial economic role (Kunin, et al, 2024). The introduction aimed to improve efficiency in production, reduce costs associated with weed control, and increase land use efficiency by enabling double cropping - growing two consecutive crops on the same land within a single year.

Economic Benefits

The economic impact of GM soybeans in Argentina has been profound. One of the most significant benefits has been the dramatic increase in soybean yields, which have grown substantially since the introduction of GM varieties. According to data, soybean production in Argentina jumped from around 11 million tonnes in the early 1990s to over 50 million tonnes in recent years. This increase is attributed largely to the adoption of GM soybeans, which allowed more extensive and intensive farming practices (Su, 2024).

Furthermore, the adoption of GM soybeans has led to a reduction in the costs associated with weed control. Farmers could use glyphosate, a relatively inexpensive and effective herbicide, over broad areas without harming the soybean crop (Brookes, 2017). This not only reduced chemical costs but also simplified crop management practices, saving labor and further enhancing profitability.

Challenges

Despite these benefits, the widespread planting of GM soybeans in Argentina has raised several challenges. One major concern has been the environmental impact, particularly the increase in herbicide use. While glyphosate is considered less harmful than many alternatives, its extensive application has led to issues like herbicide resistance in weeds, necessitating higher doses or the use of additional chemical treatments (Kanissery, et al 2019).

Additionally, the economic benefits have not been uniformly distributed. The high cost of GM seeds and the need to purchase new seeds each season (due to patent restrictions and the technology agreements) have placed a financial strain on smaller farmers. This situation has exacerbated economic disparities within the agricultural sector.

Current Status

Today, GM soybeans are deeply entrenched in Argentina's agricultural system. While they continue to provide economic benefits through high yields and efficient production, the

challenges of herbicide resistance and social disparities remain significant issues. Ongoing debates and research focus on finding sustainable farming practices that can mitigate these challenges while maintaining the economic gains provided by GM soybeans(Kunin, et al. 2024).

Conclusion

This literature review has delved into the multifaceted realm of genetically modified organisms (GMOs), offering an in-depth exploration of their development, impacts, and the promising future that lies ahead in agriculture, biotechnology, and environmental management. From the inception of recombinant DNA technology in the 1970s to the modern advances brought by CRISPR-Cas9, GMOs have evolved to become crucial tools in addressing some of the world's most pressing challenges, such as enhancing food security, improving nutritional standards, and adapting to climate change.

GMO technology has demonstrated its capacity to increase agricultural yield, reduce dependency on chemical pesticides, and fortify crops with essential nutrients. However, this technology also poses significant challenges and has sparked ongoing debates and controversies, particularly concerning safety, ethical issues, environmental impacts, and socio-economic inequalities. The case studies of Golden Rice and Bt cotton, among others, illustrate both the transformative potential and the complex dilemmas associated with GMOs.

Looking ahead, the role of GMOs in addressing global challenges like food security and climate change remains promising but requires careful navigation. Advances in genome editing have the potential to create crops that are not only more productive but also more resilient to the stresses posed by a changing climate. However, realizing these benefits demands a concerted effort to ensure that GMOs are developed and used in ways that are safe, equitable, and sustainable.

The ongoing debates surrounding GMOs highlight the need for continued research, particularly in understanding long-term ecological impacts, health effects, and the socio-economic consequences of GMO adoption. There is a critical need for transparent scientific communication and inclusive dialogue among scientists, policymakers, farmers, and the public to build trust and develop regulatory frameworks that address health and environmental safety without stifling innovation.

In contemporary society, genetically modified organisms (GMOs) are both celebrated as milestones of scientific progress and critiqued under intense public scrutiny. Their potential to contribute to sustainable development goals is clear, yet the path forward requires careful consideration of the ethical, ecological, and social dimensions, particularly when examining the differing stances of various countries, like the European Union (EU).

The EU presents a compelling case of widespread resistance to GMOs, reflecting deep-seated concerns over environmental and health impacts. European regulations on GMOs are among the strictest in the world, characterized by rigorous assessment processes, stringent labeling requirements, and a general public wariness about consuming genetically modified foods. This skepticism stems from a variety of factors including cultural values, the prioritization

of environmental sustainability, and past food safety scares that have left a lasting impact on public opinion(Lucht, J. M. 2015). In contrast to the United States, where GMOs are widely cultivated and consumed, the EU's cautious approach represents a significant resistance to fully adopting GMO technologies.

A balanced approach in evaluating the benefits and risks associated with GMOs is essential. This approach should integrate scientific rigor, ethical consideration, and public engagement to harness the benefits of GMOs while minimizing risks. Regulatory bodies and researchers must work together to ensure that GMO development aligns with broader societal goals and public well-being. In conclusion, while the road ahead for GMO technology is paved with both opportunities and challenges, its thoughtful integration into agricultural systems and biotechnological applications could play a pivotal role in shaping a sustainable and secure future for global food systems.

Works Cited

- Abbas, Mohamed S. "Genetically Engineered (Modified) Crops (Bacillus Thuringiensis Crops) and the World Controversy on Their Safety." *Egyptian Journal of Biological Pest Control*, vol. 28, no. 1, 2018, pp. 1-12, <https://doi.org/10.1186/s41938-018-0051-2>. Accessed 28 May 2024.
- Azar, Mughair A., et al. "Genetically Engineered Crops for Sustainably Enhanced Food Production Systems." *Frontiers in Plant Science*, vol. 13, 2022, <https://doi.org/10.3389/fpls.2022.1027828>. Accessed 13 Jun. 2024.
- Baranski, By: Marci, et al. "Golden Rice." *Golden Rice | Embryo Project Encyclopedia*, embryo.asu.edu/pages/golden-rice.
- Brookes, Graham. "Farm Income and Production Impacts from the Use of Genetically Modified (GM) Crop Technology 1996-2020." *GM Crops & Food*, vol. 13, no. 1, 2022, pp. 171-195, <https://doi.org/10.1080/21645698.2022.2105626>. Accessed 13 Jun. 2024.
- Brookes, Graham, et al. "The Contribution of Glyphosate to Agriculture and Potential Impact of Restrictions on Use at the Global Level." *GM Crops & Food*, vol. 8, no. 4, 2017, pp. 216-228, <https://doi.org/10.1080/21645698.2017.1390637>. Accessed 13 Jun. 2024.
- Cabelkova, Inna, et al. "The Moderating Role of Perceived Health Risks on the Acceptance of Genetically Modified Food." *Frontiers in Public Health*, vol. 11, 2023, <https://doi.org/10.3389/fpubh.2023.1275287>.
- Castro, N. G., Bjelic, J., Malhotra, G., Huang, C., & Alsaffar, S. H. (2021). Comparison of the Feasibility, Efficiency, and Safety of Genome Editing Technologies. *International Journal of Molecular Sciences*, 22(19). <https://doi.org/10.3390/ijms221910355>
- Cattaneo, Manda, et al. "Farm-scale Evaluation of the Impacts of Transgenic Cotton on Biodiversity, Pesticide Use, and Yield." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 7571, 1 June 2006, pp. 7571–6. doi:10.1073/pnas.0508312103.
- Gaines, Todd A., et al. "Mechanisms of Evolved Herbicide Resistance." *The Journal of Biological Chemistry*, vol. 295, no. 30, 2020, pp. 10307-10330, <https://doi.org/10.1074/jbc.REV120.013572>. Accessed 28 May 2024.
- Goodman, Richard E. "Twenty-eight Years of GM Food and Feed Without Harm: Why Not Accept Them?" *GM Crops & Food*, vol. 15, no. 1, 2024, pp. 40-50, <https://doi.org/10.1080/21645698.2024.2305944>.
- Kamle, Madhu, et al. "Current Perspectives on Genetically Modified Crops and Detection Methods." *3 Biotech*, vol. 7, no. 3, 2017, <https://doi.org/10.1007/s13205-017-0809-3>. Accessed 14 May 2024.
- Kanissery, Ramdas, et al. "Glyphosate: Its Environmental Persistence and Impact on Crop Health and Nutrition." *Plants*, vol. 8, no. 11, 2019, <https://doi.org/10.3390/plants8110499>. Accessed 13 Jun. 2024.
- Kathage, Jonas, and Matin Qaim. "Economic Impacts and Impact Dynamics of Bt (Bacillus Thuringiensis) Cotton in India." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 29, 2012, pp. 11652-11656, <https://doi.org/10.1073/pnas.1203647109>. Accessed 24 Jun. 2024.

- Kettenburg, Annika J., et al. "From Disagreements to Dialogue: Unpacking the Golden Rice Debate." *Sustainability Science*, vol. 13, no. 5, 2018, pp. 1469-1482, <https://doi.org/10.1007/s11625-018-0577-y>.
- Kumar, Sushil, et al. "Crop Biofortification for Iron (Fe), Zinc (Zn) and Vitamin A with Transgenic Approaches." *Heliyon*, vol. 5, no. 6, 2019, <https://doi.org/10.1016/j.heliyon.2019.e01914>.
- Kunin, Pablo Lapegna and Johana, et al. "The Impact of Soybeans in Argentina and Beyond: A Double-Edged Sword?" *ReVista*, revista.drclas.harvard.edu/the-impact-of-soybeans-in-argentina-and-beyond-a-double-edged-sword/. Accessed 13 June 2024.
- Ladics, Gregory S., et al. "Genetic Basis and Detection of Unintended Effects in Genetically Modified Crop Plants." *Transgenic Research*, vol. 24, no. 4, 2015, pp. 587-603, <https://doi.org/10.1007/s11248-015-9867-7>. Accessed 28 May 2024.
- Lucht, J. M. (2015). Public Acceptance of Plant Biotechnology and GM Crops. *Viruses*, 7(8), 4254-4281. <https://doi.org/10.3390/v7082819>.
- Nature News, Nature Publishing Group, www.nature.com/scitable/topicpage/genetically-modified-organisms-gmos-transgenic-crops-and-732/. Accessed 14 May 2024.
- Munshi, Anjana, and Vandana Sharma. "Omics and Edible Vaccines." *Omics Technologies and Bio-Engineering*, 2018, pp. 129-141, <https://doi.org/10.1016/B978-0-12-815870-8.00008-5>.
- Oliver, Melvin J. "Why We Need GMO Crops in Agriculture." *Missouri Medicine*, vol. 111, no. 6, 2014, pp. 492-507, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6173531/>. Accessed 14 May 2024.
- Peshin, Rajinder, et al. "Long-term Impact of Bt Cotton: An Empirical Evidence from North India." *Journal of Cleaner Production*, vol. 312, 1 May 2021, p. 127575. [doi:10.1016/j.jclepro.2021.127575](https://doi.org/10.1016/j.jclepro.2021.127575).
- Rodríguez-Rodríguez, D. R., Ramírez-Solís, R., Garza-Elizondo, M. A., Lourdes Garza-Rodríguez, M. D., & Barrera-Saldaña, H. A. (2019). Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). *International Journal of Molecular Medicine*, 43(4), 1559-1574. <https://doi.org/10.3892/ijmm.2019.4112>.
- Sandhu, Rubby, et al. "A Critical Review on Integrating Bio Fortification in Crops for Sustainable Agricultural Development and Nutritional Security." *Journal of Agriculture and Food Research*, vol. 14, 2023, p. 100830, <https://doi.org/10.1016/j.jafr.2023.100830>. Accessed 14 May 2024.
- Sciences, National Academies of, et al. "Social and Economic Effects of Genetically Engineered Crops." *Genetically Engineered Crops: Experiences and Prospects.*, U.S. National Library of Medicine, 17 May 2016, www.ncbi.nlm.nih.gov/books/NBK424536/. Accessed 13 June 2024.
- Siddiqui, Junaid A., et al. "Insights into Insecticide-resistance Mechanisms in Invasive Species: Challenges and Control Strategies." *Frontiers in Physiology*, vol. 13, 2022, <https://doi.org/10.3389/fphys.2022.1112278>. Accessed 28 May 2024.
- SITNFlash. "Challenging Evolution: How GMOs Can Influence Genetic Diversity." *Science in the News*, 12 Aug. 2015, sitn.hms.harvard.edu/flash/2015/challenging-evolution-how-gmos-can-influence-genetic-diversity/. Accessed 28 May 2024

New Frontiers for Sustainability in the Automotive Industry: the Case of Fuel Cells

By Fahsai Pibulsonggram

Abstract

The automotive industry has significantly contributed to environmental degradation over the years, mainly due to traditional vehicle's greenhouse gas (GHG) emissions. As the industry faces pressure to shift towards sustainability, fuel cell technology emerges as a promising frontier. This research paper will provide a comprehensive overview of fuel cell technology and explore its applications in the automotive sector. The paper will also discuss various challenges hindering the widespread adoption of fuel cell technology. Through assessing recent advancements and future prospects of fuel cell deployment in real-world settings, the paper aims to offer recommendations for fostering its development in the automotive sector. The findings are intended to inform automotive engineers, industry professionals, policymakers, environmental scientists, and academic researchers, to support informed decision-making and foster the transition towards a more sustainable future.

Introduction

Whilst the automotive industry is one key driver of economic growth and technological advancements worldwide, it has also contributed to significant environmental costs. The vast majority of automotive production focuses on traditional internal combustion engines (ICEs), which heavily rely on fossil fuels. From the early days of the Ford Model T to the modern, sophisticated vehicles of today, the automotive industry has consistently pushed the boundaries of engineering and design. As a result, the industry has greatly contributed to greenhouse gas (GHG) emissions that have led to global warming and climate change. Carbon dioxide gas (CO₂) makes up 76% of all GHG emissions (Center for Climate and Energy Solutions, 2018) and in 2020 alone, passenger cars and vans contributed to approximately one-tenth of global CO₂ emissions (Automotive Pathways to decarbonization, n.d.).

Due to the environmental consequences, there has been a growing emphasis on sustainability within the automotive industry. Many major global entities such as the EU, the US and China have set goals to achieve "net zero" or carbon neutrality within 2050 or 2060 (Automotive Pathways to decarbonization, n.d.). The push towards sustainability has led to the development and advancement in many green technologies, such as the exploration of alternative fuel sources.

Whilst achieving sustainability within the automotive industry is becoming increasingly necessary, it also has numerous barriers. Some holdbacks include the technological barrier; the large-scale adoption of fuel cell vehicles requires major investment in innovation and infrastructure development. Currently, ICEs are also more economically competitive than sustainable technologies making it a challenge to achieve a widespread adoption of these technologies (Wang, 2015). The exploration of fuel cells as a sustainable solution is promising as it offers numerous advantages that can shift the industry towards greater sustainability. Fuel cell

vehicles (FCV), have greater vehicular fuel efficiency than conventional ICEs, are simple and reliable, and produce low levels of pollution as they do not rely on fossil fuels as an energy source (Emadi and Williamson, 2004).

This research paper aims to provide a comprehensive overview of fuel cell technology, exploring the diverse applications of fuel cells within the automotive sector. It will evaluate challenges and barriers hindering the widespread adoption of fuel cell technology and assess recent advancements of fuel cell deployment in real-world settings. Moreover, the paper will offer insights into the future trajectory of fuel cell technology in the automobile industry, along with recommendations for fostering its continued development and uptake. Those who may be informed by this paper include automotive engineers, industry professionals, policymakers, environmental scientists, and academic researchers, aiming to support informed decision-making within the automotive sector and related fields. This paper will draw on literature from research papers, industry-specific sources and intergovernmental organisation sources.

Evolution of the automotive industry and its sustainability challenges

Historical reliance on ICEs

In the past century, the automotive industry heavily relied on internal combustion engine (ICE) technology due to their affordability and high energy density. An ICE generates energy through the ignition and combustion of oil-derived fuels such as gasoline or diesel (Energy.gov, 2013). The invention of the ICE began in the late 19th century, primarily developed by Nikolaus Otto and Rudolf Diesel (Environment & Society Portal, n.d.) and since then has seen continuous improvement in efficiency and reliability over the years. Technological advancements and mass production techniques such as the model T developed by Henry Ford (Ford, 2020), paved the way for the widespread adoption of ICEs. This historical moment democratised car ownership, setting the stage for the dominance of ICE technology. As a result, it embedded fossil fuel dependency into the fabric of the automotive industry and society as a whole.

Today, ICEs are still one of the most feasible options for producers and consumers within the industry, where the global market demand for ICEs reached 181,836 thousand units in 2022 and is expected to expand at a compound annual growth rate of 9.2% until 2030 (www.grandviewresearch.com, n.d.).

Environmental impacts of traditional automobile technologies

However, traditional automotive technologies such as the ICE have a high environmental cost. As the ICE relies on the combustion of carbon-based fossil fuels to generate power, it emits significant quantities of carbon dioxide (CO₂) gas. For example, the combustion of one gallon of gasoline produces approximately 8.887 kg of CO₂, and the combustion of one gallon of diesel produces approximately 10.18 kg of CO₂. Greenhouse gases (GHG) such as CO₂ play a significant role in global warming and climate change as they trap the Sun's heat within the Earth's atmosphere.

Overall, a typical passenger vehicle emits about 4.6 metric tons of carbon dioxide per year (United States Environmental Protection Agency, 2022). In 2021, passenger cars and light-duty trucks made up about 1 billion metric tons of CO₂ in the United States, making up about 16.53% of national emissions that year (Program, 2023). On a global scale, the transportation sector accounts for over one third of total CO₂ emissions, with road vehicles being the largest source (International Energy Agency, 2023).

Whilst CO₂ makes up the majority of ICE emissions, ICEs also emit other prevalent greenhouse gases including Methane (CH₄), and Nitrous oxide (N₂O). Moreover, ICEs contribute to the release of hazardous gases and particulate matter that have harmful effects on the environment and human health (Technologies, 2023). As a result, there is urgency of transitioning the automotive industry towards developing sustainable alternatives. One promising alternative to the conventional ICE is fuel cell technology.

Fuel cell technology

Fuel cells and their basic principles

Fuel cells (FC) are electrochemical devices that convert chemical energy directly into electricity using a catalyst, producing water and heat as its by-products. (Mekhilef, Saidur and Safari, 2012). Similar to an ICE, a FC uses chemical fuel such as hydrogen, oxygen, or natural gas as its energy source. However, the chemical energy is converted to electrical energy without undergoing combustion, minimising environmental harm (Haile, 2003).

Fuel cells mainly consist of an anode (negative electrode) and a cathode (positive electrode), with an electrolyte placed in between to prevent contact between the two (Universitat Politècnica de València - UPV, 2011). The global hydrogen fuel cells market has the greatest predicted growth out of all the types of fuel cells, expecting to grow to \$39.86 billion by 2032 from just \$4 billion in 2022 at a compound annual growth rate (CAGR) of 25.90% (Precedence Research, 2023). The conversion of hydrogen fuel to electrical energy can be simplified as follows: Firstly, hydrogen fuel is fed to the anode (negative electrode) where it undergoes an oxidation reaction and releases electrons:

$$\text{H}_2 (\text{g}) \rightarrow 2\text{H}^+ (\text{aq}) + 2\text{e}^-$$
At the anode, a catalyst, such as platinum or a platinum alloy, provides a surface that facilitates the breaking of hydrogen molecules into its positive ion (H⁺) and electrons, and increases the rate of reaction. The positive ions and electrons then move separately towards the cathode. Whilst the H⁺ ions move through the electrolyte, the electrons travel through an external circuit creating a flow of electricity. Electrical energy is produced from chemical energy. The H⁺ ion and electrons then undergo a reduction reaction with oxygen from the air at the cathode, producing heat and water as its product:
$$2\text{H}^+ (\text{aq}) + 2\text{e}^- + 1/2\text{O}_2 (\text{g}) \rightarrow \text{H}_2\text{O} (\text{l})$$
(Office of Energy Efficiency & Renewable Energy, 2019). Producing by-products of water and heat means that fuel cells do not create any harmful emissions and are a much more sustainable alternative to conventional energy sources.

Classifications of fuel cells and their characteristics

Overall, there are six main types of basic FCs that are appropriate energy sources for vehicles with different electrolytes and operating temperatures: proton-exchange Membrane Fuel Cell (PEMFC), Alkaline Fuel Cell (AFC), Phosphoric Acid Fuel Cell (PAFC), Molten Carbonate Fuel Cell and Solid Oxide Fuel Cell (SOFC).

Fuel cells (FC) are often classified depending on the type of electrolyte used in the cell and its operating temperature. For example, the proton-exchange membrane fuel cell (PEMFC) utilises a proton-exchange membrane as its electrolyte and operates at approximately 80°C whereas a solid oxide fuel cell (SOFC) utilises solid ceramic material as its electrolyte and operates at approximately 650–1000 °C, a much higher temperature (Pramuanjaroenkij and Kakaç, 2022). Amongst the six, the PEMFC is the most commonly used because it efficiently supports hydrogen fuel, growing in popularity as technology and infrastructure surrounding hydrogen-fueled vehicles develop. However, the SOFC is also a promising alternative as it supports a wide variety of fuels such as hydrogen, natural gas, biogas and other hydrocarbons. Each fuel cell has its unique advantages and is selected depending on the desired operating temperature, efficiency, size, and the specific requirements of the application it is intended for.

Advantages of fuel cell vehicles (FCVs) over traditional internal combustion engines (ICEs)

Within fuel cell electric vehicles (FCEV), the fuel cell stack is integrated with other components, such as the hydrogen storage tank, electric motor, and associated power electronics to make the vehicle functional and efficient. Fuel cells are often integrated into electric vehicles, either as the sole power source or as part of a hybrid system combined with rechargeable energy storage systems, such as batteries (Mo et al., 2023).

Firstly, fuel cells can be seen as a promising technology as the industry moves towards sustainability, as fuel cell vehicle emissions have a low environmental impact. Compared to conventional ICEs that emit greenhouse gases along with criteria pollutants such as PM or carbon monoxide into the atmosphere, fuel cell vehicles mainly emit water vapour and heat as key by-products reducing emission impacts on global warming, climate change or degradation in air pollution. If hydrogen fuel is used, FCVs will have zero vehicular emissions, and if hydrocarbon fuels such as are used, they will have near-zero emissions of criteria pollutants and only some carbon dioxide (CO₂) emissions depending on the selected fuel (Wang, 2002).

Moreover, fuel cell vehicles show a high energy conversion efficiency. The energy conversion efficiency of present fuel cell stacks is approximately 62% (Mo et al., 2023). When other components are taken into account, the tank-to-wheel conversion efficiency is still as high as 37.7%. This is extremely efficient compared to conventional ICEs where gasoline-powered vehicles only have an energy conversion efficiency of approximately 16% and diesel-powered vehicles have an efficiency of approximately 20%. A higher energy conversion efficiency in FCVs implies that a great portion of the fuel's chemical energy is converted into useful work to power the vehicle, and less is wasted. As a result, FCVs consume less fuel to travel the same distance as less efficient vehicles reducing emissions and fuel consumption overall. This could

also result in a lower fuel cost for fuel cell vehicles as compared to traditional vehicles. Fuel cell electric vehicles also reduce noise pollution as their components do not require combustion that produce sound, leading to a sound level difference of approximately 4-10 decibels (IEEE, 2020). As noise is a critical factor in terms of health problems such as cardiovascular disease or cognitive impairments, this reduction in acoustic impact is beneficial to human health.

Current applications of fuel cells and potential for expansion within the industry

Although there are only a few FCVs compared to ICEs or battery electric vehicles (BEVs), environmental benefits are significant. As of 2022, there were only approximately 72,000 hydrogen fuel cell vehicles (HFCV) on the world's roads as compared to 26 million BEVs (International Energy Agency, 2023) and over 1.4 Billion ICEs (Kemp, 2022). The number of HFCVs in 2022 was already 40% greater than the number of hydrogen-powered vehicles in 2021 (Collins, 2023). Whilst the market for FCV is still small, the hydrogen fuel cell vehicle market is projected to grow at a compound annual growth rate (CAGR) of 28.3% from 2023-2032 where the market size is expected to grow from \$2.8 Billion in 2022, to \$33.2 Billion in 2032. The rapid growing adoption of hydrogen fuel cell vehicles is propelled by government incentives and subsidies along with a rising public environmental consciousness.

Challenges and barriers to the adoption of fuel cell vehicles

Although fuel cell vehicles offer numerous advantages in shifting the automotive industry towards sustainability, there are still many challenges that prevent the widespread adoption of fuel cell vehicles. Firstly, the costs associated with owning a HFCV are much higher than a conventional ICE or even a battery electric vehicle. This is mainly due to the cost of the catalyst, generated from platinum group metals (PGM) that currently costs approximately \$1,000/oz and makes up 40% of the cost of a fuel cell stack (Du, Zhang and Sun, 2021). Furthermore, the energy-intensive process required for hydrogen fuel production, compression, and storage leads to a higher fuel cost for hydrogen as compared to diesel or gasoline. To be competitive with hybrid gasoline vehicles on a per-mile basis in California in 2021, hydrogen needed to have been priced at \$5.88 per kg; however, hydrogen retailed at \$16.50 per kg that year (Yowell, 2022). These high costs need to be reduced in order to make hydrogen fuel cells a feasible fuel source for the automotive industry.

The high costs of platinum group metals (PGMs) results from the scarcity of PGMs that may hinder the production of FCs. PGMs are often located in particular geographic regions: over 96% of PGMs reserves are concentrated in South Africa and Russia, which produce over 80% of PGMs in the world (Tang et al., 2023). As these regions are subject to political and economic instability, such as trade, finance and technology sanctions on Russia and labour strikes in South Africa, PGMs often experience supply disruptions (S&P Global, 2024). These disruptions reduce PGM abundance and keep prices high. There have been attempts to recycle PGMS, for example, over 150 tonnes of PGMs have been recycled per year over the last decade, 75% of which were

used for automotive catalysts (Tang et al., 2023). However, the quantity of PGMs derived from natural or secondary resources is still unable to meet their demand (Tang et al., 2023).

Whilst FCVs can reduce harmful vehicle tailpipe emissions within the automotive industry, fuel derivation processes such as hydrogen fuel extraction still create environmental impacts. Currently, more than 95% of hydrogen produced worldwide is derived from carbon-based fossil fuels, such as coal and oil (Hwang, 2013). This is mainly because it is currently cheaper to produce hydrogen fuel from GHG than through sustainable processes such as electrolysis using renewable energy sources. As a result, FCVs that run on hydrogen fuel can still contribute to GHG emissions that occur in processes before hydrogen enters the FCV. Developing sustainable hydrogen production methods such as electrolysis or using renewable energy sources faces high-cost barriers and many methods are still in research and development (Hassan et al., 2023). Public awareness that HFCVs are only partially sustainable can be a disincentive in choosing HFCVs as a sustainable option. Hydrogen is also a highly flammable gas leading to concerns with safety in storage, handling, and refuelling (Hassan et al., 2023). Although HFCVs are equipped with safety technologies, ensuring strict safety regulations and public awareness is necessary to address these concerns.

Finally, another major barrier faced by fuel cell vehicles is competition from battery electric vehicles (BEVs). BEVs benefit from better consumer awareness and more developed charging infrastructure. For example, as of January 13, 2022, there are approximately 46,290 public charging stations and 113,558 private charging outlets for EVs in the US, compared to only 43 hydrogen refuelling stations that are mostly concentrated in California (Shao & Zheng, 2023). Establishing widespread hydrogen infrastructure requires significant investments, which makes hydrogen infrastructure one of the greatest barriers to HFCV adoption. Furthermore, while fuel cell vehicles are more energy-efficient than traditional internal-combustion-engine vehicles, they are less energy-efficient than battery electric vehicles. Whilst the fuel production efficiency of an EV is up to 95%, a HFCV is only 52% efficient in fuel production due to the complex method of hydrogen production (Anon, 2023). As a result, the total energy efficiency from fuel extraction to wheels of an EV is much higher than a HFCV at 87% compared to 40%. This efficiency difference may make EVs more feasible for consumers to operate and for producers to manufacture leading to a greater adoption of EVs than FCVs.

Recent developments in the industry: case studies

Numerous nations are competing to become the leading HFCV producer fueling industry developments. There are five prominent countries that are leading in the adoption of fuel cell vehicles. In 2022, two-thirds of the additional 15,000 fuel cell cars that hit the road were in South Korea making it the leading country for the adoption of HFCV as of 2022. The government of South Korea has also set targets to have 6.2 million hydrogen vehicles on the roads by 2040 (Hassan et al., 2023). The United States holds the position as the second-largest market for FCEVs, with approximately 17,000 HFCV most of which are passenger cars. Furthermore, China holds the third-largest number of HFCVs with up to 95% of the world's hydrogen trucks

and almost 85% of the global fuel cell bus fleet, with its HFCV stock growing by 60% in 2022 (Global EV Outlook 2023, 2023b). The country aims to become a global leader in hydrogen, targeting over one million HFCVs on the roads by 2030 (Hassan et al., 2023). The Chinese government has also been heavily supporting the development of infrastructure along with firms such as BYD and Geely that have been developing and producing HFCVs.

Leading automakers' initiatives and investments in fuel cells

Major automakers such as Toyota and Hyundai have been continuously investing in their hydrogen fuel cell vehicles as a part of their sustainable strategies. As of 2024, the two existing models of HFCVs are the Toyota Mirai, which dominates HFCV manufacturing, and the Hyundai Nexo. For this year's Mirai model, Toyota minorly redesigned the exterior of their Mirai model and added additional safety features with Toyota Safety Sense 3.0, accompanying this change with a slight raise in price (Autoblog, n.d.). At the core of the Mirai model, the vehicle combines hydrogen with oxygen from the outside air to generate power without creating emissions. This mechanism has remained unchanged since the model was first released in 2014 to meet Toyota's tagline "Pure innovation. Cleaner power." (Toyota, 2023). In 2023, Toyota sold approximately 2737 Mirais, and while that number may be small, it was the best sales year ever and represented a 31% increase compared to 2022 (Car and Driver, 2023).

Whilst the Hyundai Nexo made only 241 sales, a 41% decrease compared to 2022, Hyundai has been presenting their vision for the future of hydrogen-powered vehicles. On January 8, 2024, Hyundai spotlighted its plans for utilising hydrogen energy at CES (Consumer Electronics Show) 2024 (AP News, 2024). Not only does Hyundai aim to invest more in HFCVs, but the automaker also aims to possibly aid a "hydrogen society" (AP News, 2024) by expanding into hydrogen energy production, storage and transportation through HTWO's hydrogen value chain solution. At CES, Hyundai mentioned details for the potential development of Megawatt-scale (proton-exchange membrane) PEM electrolyzer to enable green hydrogen production methods and resource circulation hydrogen production methods (HYUNDAI MOTORS, 2024). The Hyundai Motor Group projects to off-take up to 3 million tonnes of hydrogen per year by 2035. Hyundai's 2024 hydrogen-fueled Nexo has seen only minor trims from the 2018 model but is claimed to have "improved marketability by standardising customer preferences" (HYUNDAI MOTORS, 2024). The South Korean automaker claims that the selling price of the 2024 Nexo within its domestic market will be ₩69.5m (\$54,138) but is expected to be lowered to ₩37m (\$28,817) by national and local government subsidies (Martin, 2023). The South Korean government offers up to ₩22.5m (\$16,189) for new HFC passenger cars, with some cities such as Busan and Incheon offering an additional subsidy of ₩10m (\$7,195).

Another major automaker that is advancing in HFCVs is Honda with its upcoming 2025 model of the Honda CR-V e:FCEV. This model expects to "combine its plug-in charging capabilities with fast hydrogen refuelling" (Honda Automobiles, 2024) meaning that users will not only be able to refuel and hydrogen refuelling stations but also be able to charge at EV charging stations. As a lack of hydrogen refuelling infrastructure is a key barrier, this model

serves as a promising advancement for HFCVs in the automotive industry and is already beginning to gain recognition.

Recent developments in fuel cell infrastructure

Many countries are also racing to become a hydrogen leader, developing and implementing new hydrogen infrastructure to accompany hydrogen fuel cells and their applications. For example, the US government has dedicated \$7 billion to the H2 Hubs program in North America, aiming to set up 6-10 major hydrogen infrastructure around the country to serve as the foundation of a national clean hydrogen network (FASTECH, 2023). Germany has also set their goal to establish a domestic market for the production and use of hydrogen. By 2027, Germany aims to develop 1,800 kilometres of hydrogen energy pipeline network to create up to 5 gigawatts of generation capacity along with assisting European partners in their own efforts to scale up hydrogen production. As of 2023, Australia has 106 active, planned, or operational hydrogen projects accounting for approximately \$230bn–\$300bn in investment. It aims to establish a domestic hydrogen industry to supply hydrogen for various sectors such as transport, industry, and power generation by 2030. Many countries' dedication towards developing hydrogen infrastructure can lead to greater adoption of fuel cell technology as hydrogen fuel can be obtained more effectively and in higher quantities.

Advances in fuel cell materials

Over the recent years, there have been significant advancements in fuel cell catalysts. Researchers have been developing high-performing catalysts to boost the reaction rates within the fuel cell to increase efficiency and reduce costs (FASTECH,2023). To achieve this, research has focused on developing platinum-alloy catalysts. Scientists have found that certain platinum alloys with larger surface area have the ability to enhance oxygen diffusion and proton surface conduction (Shao and Ni, 2024). Overall, recent research and development has improved the performance of specific platinum-group-metal-free catalysts by approximately 60% from the 2021 baseline to 2023 (Storage, 2023). To improve the durability of fuel cell catalysts, researchers have also begun to develop graphene-based catalysts that show 30% less activity loss in rigorous stress tests, setting a new standard for durability in fuel cell technology (FASTECH, 2023).

Furthermore, one fuel cell that researchers have been constantly developing is the proton-exchange membrane fuel cell (PEMFC). Researchers are working to improve chemical, mechanical, and thermal stability to allow the PEMFC to adapt to specific applications, such as dry conditions or higher operating temperatures. (Shao and Ni, 2024). Through developing thinner electrolyte membranes that reduce ohmic losses by shortening the proton and water transport path, PEMFCs have seen improved durability for operating at high temperatures over 100°C. As a result, PEMFCs serve as a promising technology for enhancing the overall efficiency and reliability of fuel cell systems. The integration of fuel cells with other components, such as batteries in hybrid systems has also been seen to increase the overall energy

efficiency of fuel cells (FASTECH,2023). Thus, fuel cell hybrids are also a promising technology for scientists to explore further.

Fuel cell employment examples in public transportation and beyond

Hydrogen fuel cells are slowly being integrated into the public transportation sector, where London has successfully deployed 20 hydrogen-powered double-decker buses to accelerate the city's goal of reaching zero emissions (Nhede, 2021). The local government body Transport for London provided up to £6 million in funding for this project with a goal to help improve air quality and health for Londoners by reducing the level of harmful nitrogen oxide in the air. This project was part of a joint initiative for hydrogen-powered vehicles across Europe (JIVE) that seeks to deploy 139 new fuel cell buses across five European countries. Within the transportation industry as a whole, there is also progress of HFC adoption in the shipping industry. For example, the Norwegian shipyard Myklebust Verft has been tasked with building two hydrogen-powered ferries that will be deployed in the year 2026 (Prevljak, 2024). These ferries will be the largest hydrogen-powered ships in the world and are expected to reduce CO₂ emissions by approximately 26,500 tons each year, running 85% on hydrogen. These examples provide a promising outlook for greater adoption of hydrogen fuel cell technology within the automotive and transportation industry.

Applications of fuel cells expand beyond the transportation industry. Within the healthcare industry, a company called Bloom Energy has provided fuel cell systems for seven healthcare facilities that offset between 37% and 81% of their utility consumption (Mazzetti, 2016). The Bloom Box Beta version used is a 200 kW solid oxide fuel cell (SOFC) that utilises natural gas or biogas. Whilst this project costs up to \$62.057M, the project with Bloom Energy is anticipated to offset over 15 million lbs of CO₂. Hydrogen fuel cells are also being used in industrial applications. For example, Amazon and Walmart have begun to use hydrogen fuel cell forklifts in their warehouses as they reduce emissions and also offer longer operation times and faster refuelling than battery-powered alternatives (Wordsworth, 2024). This shows that fuel cell technology can be a sustainable alternative across various industries and fuel cell adoption as a whole will slowly expand into different areas and uses.

Future outlooks and recommendations

Fuel cell technology serves as a promising frontier for sustainability across various industries. Within the automotive industry, hydrogen fuel cell vehicles are increasingly integrated into major automakers' plans for achieving sustainability, and are expected to expand and grow in adoption in the future. However, whilst FCVs clearly reduce harmful emissions compared to conventional ICES, further research and development must be done to truly reduce environmental impacts as most hydrogen production methods today are still produced from fossil fuels. Whilst fuel cell technology has great potential, the widespread adoption of electric vehicles has also overshadowed other sustainable alternatives in the automotive industry. EVs are accompanied by sufficient infrastructure, lower cost and greater consumer awareness. As a

result, it will be challenging for FCVs to become a leading green alternative. At the same time, FCVs will not slowly disappear from the industry but, instead, continue to grow alongside EVs until sufficient development makes operating an FCV as efficient and practical as an EV.

To ensure the widespread adoption of fuel cell technologies, future fuel cell research should focus on developing factors that lower production costs, improve efficiency and ensure zero carbon. For example, to lower production costs, researchers should focus on methods to develop less costly catalysts or explore catalysts that do not showcase the same need for rare, costly materials. Researchers should also look into new fuel cell materials such as graphene or the development of PEMFCs that can help lower costs, and have potential in increasing the capabilities of fuel cells in different circumstances, such as high temperatures. Additionally, future research could investigate methods to reduce the energy intensity required for hydrogen fuel compression and storage to reduce its costs of production. One of the most crucial areas of research should focus on the development of green hydrogen production methods. Whilst grey hydrogen, produced from fossil fuels, dominates the current hydrogen market, green hydrogen offers a fully sustainable alternative. It is produced by splitting water by electrolysis which produces only hydrogen and oxygen (Haynes, 2022). Whilst this process requires great investments, the advancement of green hydrogen will allow hydrogen fuel cells (HFCs) to gain competitiveness as a sustainable alternative for the automotive industry and other industries, with significant potential for reducing emissions and environmental impact.

To support this transition, policymakers should continue to implement policies that focus on reducing tailpipe emissions, such as subsidising or reducing taxes for HFCVs. Additionally, governments should begin to invest more in green hydrogen production technologies and hydrogen charging infrastructure. Although these investments are substantial, they are necessary in transforming fuel cell technology into a completely sustainable alternative that can help shift the automotive industry and various other industries to reduce their emissions and environmental impacts.

Beyond the automotive industry, fuel cell technology has a high possibility of rapid expansion and adoption. Whilst one certain industry may not become completely hydrogen-powered, as environmental consciousness increases, many industries will begin to experiment and explore the possibilities of fuel cell technology, as we have recently seen in logistics and healthcare industries. Furthermore, as fuel cell technologies develop, more industries will be able to easily adopt and integrate fuel cell technology to reduce their environmental impacts.

Works Cited

- Anon, (2023). Energy Efficiencies: EVs Versus FCVs - ENERGYminute. [online] Available at: <https://energyminute.ca/infographics/energy-efficiencies-evs-versus-fcvs/>.
- AP News. (2024). Hydrogen energy back in the vehicle conversation at CES 2024. Available at: <https://apnews.com/article/ces-2024-car-technology-unveils-7d825a33d810a9b29ebf260404fe81cf>.
- AtoB Fuel Card. (n.d.). ICE vs EV Vehicles: Effects on the Environment | AtoB. [online] Available at: <https://www.atob.com/blog/ice-vs-ev-vehicles-effects-on-the-environment>.
- Autoblog. (n.d.). 2024 Toyota Mirai gets tiny changes, small price increases. [online] Available at: <https://www.autoblog.com/2024/01/27/2024-toyota-mirai-updates-price-increase-safety-features/> [Accessed 20 Jul. 2024].
- Collins, L. (2023). The number of hydrogen fuel-cell vehicles on the world's roads grew by 40% in 2022, says IEA report. [online] Hydrogen news and intelligence | Hydrogen Insight. Available at: <https://www.hydrogeninsight.com/transport/the-number-of-hydrogen-fuel-cell-vehicles-on-the-worlds-road-s-grew-by-40-in-2022-says-iea-report/2-1-1444069>.
- Du, L., Zhang, G. and Sun, S. (2021). Proton Exchange Membrane (PEM) Fuel Cells with Platinum Group Metal (PGM)-Free Cathode. *Automotive Innovation*, 4(2), pp.131–143. doi:<https://doi.org/10.1007/s42154-021-00146-0>.
- Energy.gov. (2013). Internal Combustion Engine Basics. [online] Available at: <https://www.energy.gov/eere/vehicles/articles/internal-combustion-engine-basics>.
- Environment & Society Portal. (n.d.). Invention of the Internal Combustion Engine. [online] Available at: <https://www.environmentandsociety.org/tools/keywords/invention-internal-combustion-engine#:~:text=The%20inventions%20could%20power%20automobiles>.
- Ford (2020). The Model T. [online] Ford Corporate. Available at: <https://corporate.ford.com/articles/history/the-model-t.html#:~:text=The%20Model%20T%20was%20introduced>.
- Global EV Outlook 2023. (2023a). Available at: https://iea.blob.core.windows.net/assets/525aa16b-7a9d-40f9-a89f-5e613f019220/GEVO2023_WEB.pdf.
- Global EV Outlook 2023. (2023b). Available at: https://iea.blob.core.windows.net/assets/525aa16b-7a9d-40f9-a89f-5e613f019220/GEVO2023_WEB.pdf#:~:text=URL%3A%20https%3A%2F%2Fiea.blob.core.windows.net%2Fassets%2F525aa16b [Accessed 9 Jul. 2024].
- Haile, S.M. (2003). Fuel cell materials and components. *Acta Materialia*, 51(19), pp.5981–6000. doi:<https://doi.org/10.1016/j.actamat.2003.08.004>.
- Hassan, Q., Azzawi, I.D.J., Sameen, A.Z. and Salman, H.M. (2023a). Hydrogen Fuel Cell Vehicles: Opportunities and Challenges. *Sustainability*, [online] 15(15), p.11501. doi:<https://doi.org/10.3390/su151511501>.
- Hassan, Q., Azzawi, I.D.J., Sameen, A.Z. and Salman, H.M. (2023b). Hydrogen Fuel Cell Vehicles: Opportunities and Challenges. *Sustainability*, [online] 15(15), p.11501. doi:<https://doi.org/10.3390/su151511501>.
- Haynes, A. (2022). The difference between green hydrogen and blue hydrogen. [online] Petrofac. Available at: <https://www.petrofac.com/media/stories-and-opinion/the-difference-between-green-hydrogen-and-blue-hydrogen/>.

- Holtze, B. (n.d.). World's first hydrogen-powered ship is powered by Ballard fuel cells. [online] blog.ballard.com. Available at: <https://blog.ballard.com/marine/worlds-first-liquid-powered-hydrogen-ship-mf-hydra-is-powered-by-ballards-fuel-cells>.
- Honda Automobiles. (2024). 2025 Honda CR-V FCEV 1 Eco-Conscious Powerhouse 1 Honda. [online] Available at: <https://automobiles.honda.com/future-cars/cr-v-fcev>.
- Hwang, J.-J. (2013). Sustainability study of hydrogen pathways for fuel cell vehicle applications. *Renewable and Sustainable Energy Reviews*, [online] 19, pp.220–229. doi:<https://doi.org/10.1016/j.rser.2012.11.033>.
- HYUNDAI MOTORS. (n.d.). Hyundai Motor Reveals Vision for Hydrogen Energy and Software Solutions Beyond Mobility at CES 2024. [online] Available at: <https://www.hyundai.com/worldwide/en/newsroom/detail/hyundai-motor-reveals-vision-for-hydrogen-energy-and-software-solutions-beyond-mobility-at-ces-2024-0000000394>.
- International Energy Agency (2023a). Transport. [online] IEA. Available at: <https://www.iea.org/energy-system/transport>.
- International Energy Agency (2023b). Trends in electric light-duty vehicles. [online] IEA. Available at: <https://www.iea.org/reports/global-ev-outlook-2023/trends-in-electric-light-duty-vehicles>.
- Kemp, A. (2022). How Many Cars Are There in the World? [online] PD Insurance. Available at: <https://www.pd.com.au/blogs/how-many-cars-in-the-world/#:~:text=As%20of%202022%2C%20there%20are>.
- Luo, Y., Wu, Y., Li, B., Mo, T., Li, Y., Feng, S.-P., Qu, J. and Chu, P.K. (2021). Development and application of fuel cells in the automobile industry. *Journal of Energy Storage*, 42, p.103124. doi:<https://doi.org/10.1016/j.est.2021.103124>.
- Martin (64635f1b41c62), P. (2023). Hyundai unveils the 2024 version of its hydrogen-powered Nexo fuel-cell car. [online] Hydrogen news and intelligence | Hydrogen Insight. Available at: <https://www.hydrogeninsight.com/transport/hyundai-unveils-the-2024-version-of-its-hydrogen-powered-nexo-fuel-cell-car/2-1-1465779>.
- Mo, T., Li, Y. and Luo, Y. (2023a). Advantages and Technological Progress of Hydrogen Fuel Cell Vehicles. *World Electric Vehicle Journal*, [online] 14(6), p.162. doi:<https://doi.org/10.3390/wevj14060162>.
- Mo, T., Li, Y. and Luo, Y. (2023b). Advantages and Technological Progress of Hydrogen Fuel Cell Vehicles. *World Electric Vehicle Journal*, [online] 14(6), p.162. doi:<https://doi.org/10.3390/wevj14060162>.
- Nhede, N. (2021). London launches England's first hydrogen-powered double-decker buses. [online] Smart Energy International. Available at: <https://www.smart-energy.com/regional-news/europe-uk/london-launches-englands-first-hydrogen-powered-double-decker-buses/> [Accessed 25 Jul. 2024].
- O'Connor, C. and Alexandrova, T. (2021). The Geological Occurrence, Mineralogy, and Processing by Flotation of Platinum Group Minerals (PGMs) in South Africa and Russia. *Minerals*, 11(1), p.54. doi:<https://doi.org/10.3390/min11010054>.
- Office of Energy Efficiency & Renewable Energy (2019). Fuel Cells. [online] Energy.gov. Available at: <https://www.energy.gov/eere/fuelcells/fuel-cells#:~:text=A%20fuel%2C%20such%20as%20hydrogen>.

- Pramuanjaroenkij, A. and Kakaç, S. (2022). The fuel cell electric vehicles: The highlight review. *International Journal of Hydrogen Energy*. doi:<https://doi.org/10.1016/j.ijhydene.2022.11.103>.
- Precedence Research (2023). Hydrogen Fuel Cells Market Size, Growth, Trends, Report 2023-2032. [online] [Precedenceresearch.com](https://www.precedenceresearch.com). Available at: <https://www.precedenceresearch.com/hydrogen-fuel-cells-market#:~:text=The%20global%20hydrogen%20fuel%20cells%20market%20size%20is%20expected%20to> [Accessed 30 Aug. 2024].
- Prevljak, N.H. (2024). Norwegian shipyard to build the ‘world’s largest’ hydrogen-powered ships. [online] *Offshore Energy*. Available at: <https://www.offshore-energy.biz/norwegian-shipyard-to-build-the-worlds-largest-hydrogen-ships/> [Accessed 25 Jul. 2024].
- Program, U.S.G.C.R. (2023). Fifth National Climate Assessment. [online] nca2023.globalchange.gov. Available at: <https://nca2023.globalchange.gov/chapter/13/>.
- Shao, Z. and Ni, M. (2024). Fuel cells: Materials needs and advances. *MRS bulletin*. doi:<https://doi.org/10.1557/s43577-024-00722-9>.
- Statista (2023). Statista - The Statistics Portal. [online] *Statista*. Available at: <https://www.statista.com/aboutus/our-research-commitment>.
- Storage, O. (n.d.). Hydrogen and Fuel Cells Advancing Technologies for Producing, Delivering, and Storing Hydrogen PRODUCTION Accelerated hydrogen production R&D. Achieved a world record in solar-to-hydrogen conversion efficiency in a photoelectrochemical device and screened more than 1,000 materials for thermochemical hydrogen production. FUEL CELL R&D HYDROGEN R&D System (heavy-duty) Production (electrolytic hydrogen). [online] Available at: <https://www.energy.gov/eere/fuelcells/articles/progress-hydrogen-and-fuel-cells>.
- Tang, H., Peng, Z., Tian, R., Ye, L., Zhang, J., Rao, M. and Li, G. (2023). Platinum-group metals: Demand, supply, applications and their recycling from spent automotive catalysts. *Journal of Environmental Chemical Engineering*, 11(5), pp.110237–110237. doi:<https://doi.org/10.1016/j.jece.2023.110237>.
- Technologies, Delta-Q. (2023). The Effects of Internal Combustion Engines on the Environment. [online] *Delta-Q Technologies*. Available at: <https://delta-q.com/industry-news/the-effects-of-internal-combustion-engines-on-the-environment/#:~:text=The%20pollution%20caused%20by%20ICEs>.
- Toyota (2023). 2021 Toyota Mirai Fuel Cell Vehicle | Innovation is Power. [online] www.toyota.com. Available at: <https://www.toyota.com/mirai/>.
- United States Environmental Protection Agency (2022). Greenhouse Gas Emissions from a Typical Passenger Vehicle. [online] *US EPA*. Available at: <https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle>.
- Universitat Politècnica de València - UPV (2011). Main Components of a Fuel Cell | 3/14 | UPV. [online] *YouTube*. Available at: <https://youtu.be/TpJm9AGrc4U?si=DmKv-5byOksfrLxK> [Accessed 4 Jul. 2024].
- Voelcker, J. (2022). Hydrogen Fuel-Cell Vehicles: Everything You Need to Know. [online] *Car and Driver*. Available at: <https://www.caranddriver.com/features/a41103863/hydrogen-cars-fcev/>.

- Wang, M. (2002). Fuel choices for fuel-cell vehicles: well-to-wheels energy and emission impacts. *Journal of Power Sources*, 112(1), pp.307–321.
doi:[https://doi.org/10.1016/s0378-7753\(02\)00447-0](https://doi.org/10.1016/s0378-7753(02)00447-0).
- Wordsworth, S. (2024). Amazon installs hydrogen production system for forklifts. [online] *Industrial Vehicle Technology International*. Available at:
<https://www.ivtinternational.com/news/hydrogen/amazon-installs-hydrogen-production-system-for-forklifts.html> [Accessed 25 Jul. 2024].
- Wulandari, F. (2024). [online] *Techopedia*. Available at:
<https://www.techopedia.com/investing/platinum-price-forecast> [Accessed 1BC].
- www.expertmarketresearch.com. (n.d.). Global Green Hydrogen Market Report and Forecast 2024-2032. [online] Available at:
<https://www.expertmarketresearch.com/reports/green-hydrogen-market> [Accessed 3 Jul. 2024].
- www.fastechus.com. (n.d.). Innovations in Hydrogen Fuel Cell Technology: Exploring the Latest Developments | FASTECH. [online] Available at:
<https://www.fastechus.com/blog/innovations-in-hydrogen-fuel-cell-technology>.
- www.fastechus.com. (n.d.). The Global Race To Establish Hydrogen Hubs | FASTECH. [online] Available at: <https://www.fastechus.com/blog/hydrogen-hubs> [Accessed 23 Jul. 2024].
- www.fortunebusinessinsights.com. (n.d.). Automotive Fuel Cell Market Size, Share & Insights by 2028. [online] Available at:
<https://www.fortunebusinessinsights.com/industry-reports/automotive-fuel-cell-market-101809>.
- www.google.com. (n.d.). Redirecting. [online] Available at:
<https://www.google.com/url?q=https://www.sciencedirect.com/science/article/abs/pii/S0360544220320624&sa=D&source=docs&ust=1720501276958459&usg=AOvVaw0JJAjGAVtqICTVJtPJ4UWn> [Accessed 9 Jul. 2024].
- www.grandviewresearch.com. (n.d.). Internal Combustion Engine Market Size & Share | Industry Report, 2025. [online] Available at:
<https://www.grandviewresearch.com/industry-analysis/internal-combustion-engine-market>.
- www.mazzetti.com. (2016). Healthcare On-Site Renewable Energy (Fuel Cells) | Mazzetti. [online] Available at: <https://www.mazzetti.com/project/bloom-energy-fuel-cells/> [Accessed 25 Jul. 2024].
- www.plugpower.com. (2024). Hydrogen Fuel Cell Forklifts: The Impact of Fuel Cells in Material Handling - Plug Power. [online] Available at:
<https://www.plugpower.com/hydrogen-fuel-cell-forklifts-the-impact-of-fuel-cells-in-material-handling/#:~:text=Not%20only%20do%20hydrogen%20fuel> [Accessed 13 Jul. 2024].
- www.spglobal.com. (2024). Sanctions against Russia – a timeline. [online] Available at:
<https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/sanctions-against-russia-8211-a-timeline-69602559>.
- Yowell, G. (2022). How does the cost of Hydrogen stack up against gasoline? [online] *Stillwater Associates*. Available at:
<https://stillwaterassociates.com/how-does-the-cost-of-hydrogen-stack-up-against-gasoline>

In Which Ways do Superblocks Reflect on and Integrate with the Fundamental Values of the Local Mediterranean/Spanish cultures? By Lucas Srba Cabeza

Abstract

This project explores the forward-thinking urban planning concept known as “Superblocks” in Barcelona, Spain. This work builds upon the author’s own research conducted in Barcelona along with his experiences as a resident of this city over the past ten years. Particular aspects of the Superblocks that are of interest for this paper are: mobility and transport; the natural ecosystem; air quality and public health; familial and communal culture; gentrification issues, and the aesthetic and psychological considerations behind good public spaces. This paper primarily looks at how the Superblocks project’s genesis, development, and implementation reflect and integrate with the Catalan/Mediterranean culture of Barcelona’s residents. Areas of exploration will include the history of the city, its different eras of planning and design and governance, residents’ attitudes toward their neighbors, leisure time, the environment, their rituals and habits. From this analysis, a determination will be made about the appropriateness of the Superblocks project for Barcelona and its citizens, the extent to which it is or will be successful, and whether it is a model that can be applied in other cities. Given the available data presented in this paper, it is anticipated that Superblocks, indeed, reflect on and integrate with the fundamental values of the local Mediterranean/Spanish cultures.

Context

The way I got to really know about Barcelona was the way I suggest anyone should learn about a city: avoid tourist spots, go straight to the center, and immerse yourself in the neighborhoods, parks, and local streets. You’ll get a good feel for what actually makes a city breathe, and if you are lucky, you might just learn some things about the people who live there.

I have made this principle my guiding light during the past decade that I’ve lived in Barcelona. Time and again, I’ve gotten good at getting lost right in my own backyard. Even better, I still lose myself in many of the same neighborhoods and streets where I’ve been wandering this whole time. As my understanding of how neighborhoods and cities evolves, I have developed a deep love of Barcelona. Or at a minimum, I have become an observant student of Barcelona’s seasons and changes.

That is why, a few years ago while walking around the renovated Mercat de Sant Antoni, something really grabbed my attention. In my journeys over the years, I had noticed work being done here, and now that it was finished it was just like when the pieces of a jigsaw puzzle all rush into place.

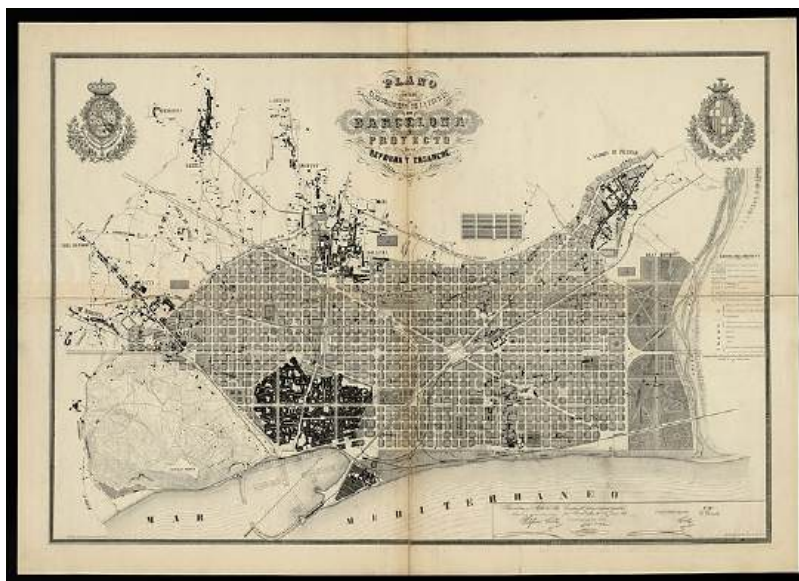
All around the market, and for as far as I could see, the busy streets were transformed into tree-lined pedestrian spaces. The cars that had previously sliced up the streets were now gone, replaced with flowers and greenery. On the newly installed benches and tables, elderly couples chatted in the sun while children played impromptu games. Unusually for a city as hectic

as Barcelona, the whole area had a sense of calm about it. It was unlike anything I'd ever seen or experienced in all of my years exploring the streets of this city.

I wanted to know more about what had happened here and how this transformation had taken place. When I got home, I jumped online and in seconds I discovered that what was happening was way bigger than just what was going on in the area around the market of Sant Antoni. It was a revolutionary idea, but it really shouldn't have surprised me. Let me tell you why.

Barcelona is an ancient city. By the time the Romans settled the town of Barcino almost 2,000 years ago, it had already been occupied by both Iberians and Carthaginians or Phoenecians (de la Torre-Moral et al.). But more relevant to the current changes in the city's urban fabric is the more recent history of Barcelona's innovative urban planning. In the 1850s, Barcelona was a city of almost 200,000 people still largely contained within the original city walls. (The Guardian) Due to overcrowding, garbage, human waste, and disease were rampant. The city needed to expand beyond its walls; however, there was some opposition to change as military commanders pointed out that the walls had protected the city well over the years. (Wynn) Eventually, the city council prevailed and the commission for the design of the expanded city was awarded to a plan developed by Catalan engineer Ildefons Cerdà. This came to be known as the Plan Cerdà. (Urbano)

The Plan Cerdà sought to solve Barcelona's overcrowding by creating streets laid out in a regular grid pattern, connecting the old city with seven outlying neighborhoods. At a very fundamental level, this plan improved living conditions for residents and established a strong foundation for future development. Cerdà's original plan created empty spaces in the center of each city block to prioritize quality human living conditions like plentiful light, air, and an abundance of greenery. The wide avenues allowed different forms of transport to use the space as well.



Cerdà's plan for the Eixample. Image credit: Ildefons Cerdà, found on Barcelonayellow.

This new development suited the burgeoning Mediterranean culture in Barcelona. By allowing residents to walk from place to place without the constraint of city walls, the possibility of interacting with neighbors, shopkeepers, and friends along the way might logically increase. The neatly laid-out streets allowed for a healthy mix of apartments, shops, businesses, restaurants, and services all nestled together, often creating micro-cities within each city block. (Tatjer) Even before the dawn of the motor vehicle, horses and bicycles could easily move people and goods from place to place, laying the groundwork for Barcelona's strong public transport system. (TMB)

Over 100 years later, Barcelona responded to the rapid urbanization of the late 20th century with a new urban planning concept. Cars had exerted a chokehold on the city, and green space was at an all time low. (Ariadna Baró) Noise and air pollution was responsible for 350 additional deaths per year. (Ajuntament de Barcelona [1]) In response to these issues, in 1987, urban planner Salvador Rueda proposed an inspired public works project. With ambitions to remove cars from the city, provide a cleaner environment, and create several tree-filled "green axes" for residents to enjoy, his plan aimed to resolve some of the major problems that Barcelona faced. He called this innovative plan "Superblocks", (Roberts [1]) the first of which was created in the el Born district of Barcelona in 1993. After a period of relative quiet, Superblocks entered their "golden age" of growth in the decade since the introduction of the 2013-2018 Urban Mobility Plan of Barcelona. To date, five full Superblock projects have been completed in the Eixample district of the city.

Literature Review

Much has been said about the proliferation of Superblocks in Barcelona over the last 10 years. Architects, planners, and scholars have both praised their transformative power and derided them over the negative effects of gentrification and housing equity. In reviewing the academic and critical literature written about Superblocks in Barcelona, four key themes present themselves. First, since modern superblocks arose out of the multi-purpose 2013-2018 Urban Mobility Plan for Barcelona (Ajuntament de Barcelona [2]), topics related to traffic and pollution in the city center appear frequently. Secondly, consistent with Barcelona's intentions to create a healthier climate for residents, much research covers the health and psychosocial effects of green spaces and pedestrian-friendly areas within the city center. Third, much has been written on the process involved in going from typical city thoroughfares to a more human-centered and walkable city. Finally, significant literature exists on issues of gentrification, housing equity, and potential discrimination within Superblocks themselves. An overriding theme that presents itself in the literature is the human-centered impact in the wake of the creation of Superblocks that benefits the populace who actually live in the center of the city. Interestingly, even though former Mayor Ada Colau stated "omplim de vida els carrers" ("Let's fill these streets with life") in her 2016 post-election speech, very little has been written about the impact of Superblocks on traditional cultures and values of Spanish and Mediterranean life.

In the face of this largely unexamined topic, this paper intends to explore how Superblocks impact the traditional values and culture of this Mediterranean Spanish community. Is the goal of returning the city to the locals actually consistent with the traditional values of the community? In order to consider this question, I undertook a survey of literature relating to the cultural preferences and workings of cities in the northern Mediterranean in general. As a complement to this research, I conducted independent interviews with two key figures integral to the modern Superblock movement in Barcelona. The first was with local planner Salvador Rueda who is the founder of the Urban Ecology Agency. He is often credited with first describing the concept of Superblocks back in 1987. He has long been a proponent of the implementation of these plans in Barcelona, with the motivation being to return the city center back to the residents. The second interview was conducted with Xavier Matilla, former chief architect of the city of Barcelona from 2019-2023. The interview focused on his experience with translating an architectural and planning concept consistent with the lifestyles of the residents with the realities of a vigorous political landscape.

Analysis

Mediterranean culture is uniquely built around the importance of family. According to Laudani, et al., “In Mediterranean countries, the term family indicates a diffuse network of relationships, obligations, and loyalties.” Woven through both the consequential and quotidian events of life in these regions are the family references that serve as guideposts for each person. Whether they are sprawling or small, families and interactions between them are important in Mediterranean countries. In Spain, it is not unusual for three or four generations of family to live under the same roof, eating and socializing together every day. For families living in separate dwellings, it would be a rarity if they did not come together for a meal or an event at least once per week (de la Torre-Moral et al.).

Family life in Barcelona may well have influenced the decisions of the planners and architects who first envisioned the idea of Superblocks. But has the reality of Superblocks lived up to the goal of improving the lives of families in the city center? On the face of it, the answer seems to be ‘yes,’ assuming that in Mediterranean cultures, merely being together is a good starting point for families and extended families. Superblock planners have taken into consideration many elements conducive to family bonding time, and tried to build upon them. This applies equally for younger people, who have been provided with an abundance of play spaces, picnic tables, and plazas, and also for the grandparents, who have benches where they can sit, relax, and enjoy the sunshine. According to the Ajuntament (Town Hall) of Barcelona, “Nearly 1,000 new items of furniture (benches, seats, children's games) and new lighting” will be installed and nearly a ten-fold increase in greenery will be seen. (Ajuntament de Barcelona [3]) As an escape from an otherwise cramped Eixample apartment, Superblocks provide common areas and ample opportunities for families to spend time together.

Along with family, community is an integral part of life for residents of Mediterranean cities. Over time, densely populated and historic cities like Barcelona have come to highly value

the importance of community. In Barcelona's case, the concept of working and living together was an important element of simply surviving in a city that faced numerous attacks over the past 15 centuries. (Museu d'Història de Catalunya) In contemporary times, however, an appreciation of community has developed as a rewarding way for residents to build a sense of belonging, bonding, and commitment to those who live and work around them. The generally pleasant climate in Mediterranean cities like Barcelona also facilitates outdoor interactions between those living in the community.

One illustration of the way Superblocks have enriched community life in Barcelona is by observing the tradition of large, planned outdoor parties in the city. In towns and cities all over Spain, residents dearly value the much anticipated 'neighborhood parties' (Fiestas del Barrio in Spanish): occasions where neighbors come together to complete events and activities, sometimes over the course of many days. (Ajuntament de Barcelona [4]) They are held outdoors and local participation rates are high as residents are drawn to the shopping stands, music performances, food stalls, traditional dances, and religious festivals. In the center of Barcelona, before the introduction of Superblocks, typically only small spaces could be blocked off by the city for these Fiestas. At times, this resulted in crowded and sometimes shortened events as the residents of these thickly populated areas could not comfortably fit in the spaces set aside for the community parties. With the advent of Superblocks in Barcelona, this has now changed. One of the biggest community parties of the year is in the Barrio (Neighborhood) of Sant Antoni where the central market has been recently renovated, with blocks of Superblocks now surrounding the area. Thanks to these changes, the local residents (and now, many visitors) are able to use the newly created open spaces they need for their Fiesta del Barrio.

Unlike shopping patterns and preferences in much of the United States, residents of Mediterranean cities prefer to do their shopping in a more personal and localized manner. (Blanchar) Mediterranean shopping is generally a small-scale, daily experience. Merchants are seen as community members and neighbors, and it is not uncommon for a fellowship to develop between local shoppers and the vendors. Centralized shopping patterns are common in Barcelona. For example, each distinct neighborhood of the city has its own open-air market where foodstuffs and goods of all types are sold in variety and abundance. In total, there are 39 such markets in the city (Meet Barcelona) and irrespective of one's shopping preferences, long-standing tradition dictates that citizens walk to the market with their preferred wheeled carry carts to buy the basics of their meals. In a sense, markets represent a centuries-long tradition of tying city communities to agricultural areas while also reaffirming an individual's belonging to a cycle of life experiences that cement one's place in the community.

Given this backdrop of how residents shop in Mediterranean communities, it seems logical to observe that the implementation of Superblocks has enhanced the shopping experience for all. First, for the merchants: Their volumes of business have increased as it has become simpler for residents and non-residents alike to reach their shops (Matilla). Concerns about parking or worries about walking through heavy traffic have lessened. While residents of the neighborhoods may still prefer to shop locally, Superblocks have provided easier and safer ways

for them to utilize stores that are slightly farther away without having to resort to any form of public transit. Pavements are smoother than when there were formerly streets. Because of this, residents are able to transport their purchases in wheeled carts, often without even having to stop for intersections at busy streets. In sum, the shopping experience with Superblocks in place has become an improved experience for residents while respecting and enhancing the fundamental Mediterranean shopping ethos of centrality, locality, and ease of access.

Mediterranean cities have developed over centuries with certain advantages and challenges in terms of transportation. Many of these communities have built expansive waterway transport systems over the millennia, fostering a culture of trading with far-away people within the Mediterranean and as far away as the Americas. Within the cities, however, transport preferences are driven by the relatively pleasant weather and proximity of living spaces to businesses, shops, work, and places of worship. Walking evolved as an efficient method for transportation since the earliest days of cities as walled fortresses. The complications involved with navigating transport options like horses, wagons, or cars in a compact and densely populated space may well have solidified residents' preferences for walking. In recent years, Barcelona's residents have embraced bicycles as a convenient form of transport, and the city has installed short-term bike rentals in virtually every neighborhood (SMOU). They have given residents and visitors greater flexibility in navigating the city in a simple and convenient manner.

Superblocks have enhanced the Mediterranean penchant for walking and cycling. Wider sidewalks and converted roadways offer residents greater ownership of the areas outside of their homes in a way that was not possible when the entire city was open to automobile traffic. Walking is still preferred by many and the enhanced shade canopy (planted during the conversion to Superblocks) affords more shade during hot summer months. Conversely, in the center of the former streets, the canopy is mostly open, allowing residents to enjoy the warmth of the sun during the winter months. Stretches of uninterrupted spaces that came with the advent of Superblocks have provided easier access to metro and bus stops and have eliminated the need to cross busy roads in many places. Most critically, however, Superblocks were born under a grand Urban Mobility Plan that was specifically designed to make transport work better within the city of Barcelona. Almost 300 kilometers of dedicated bicycle lanes have been put into place as part of this master plan for Superblocks as of 2023 (Ajuntament de Barcelona [5]).

For residents of Mediterranean communities, living in a healthy and balanced manner is an integral part of life. In fact, Spain was recently named the healthiest country in the world. (EL PAÍS) Some manner of exercise, combined with a fresh and healthy diet, are readily observable in cities like Barcelona. Residents organically find ways to incorporate this mentality into their lives each day, unlike in many other cultures where residents are forced to schedule in an hour of exercise in a gym before rushing off to work. With over 5,000 sports installations for residents of Barcelona, there is always a workout facility nearby (Generalitat de Catalunya). Walking in Mediterranean cities has achieved ritual status in the form of the daily paseo (or "stroll" in English) where residents simply take a walk in the late afternoon or evening hours. In addition to the exercise of walking, during these paseos many important elements of Mediterranean life are

combined as they are often taken with family or friends, pass local community members or merchants, and combine exercise with taking a leisurely meal outside. This has been the way of life for so long that citizens intuitively understand the importance of such a lifestyle for their health and well-being. (Muñoz)

Superblocks have provided new venues for citizens of Barcelona to realize a healthy lifestyle with plenty of exercise. Long, contiguous blocks of former streets created by Superblocks make for a magnified space for walks or rides. On the street of Consell de Cent, for example, strollers can walk 20 city blocks (Ajuntament de Barcelona [6]) without being interrupted by traffic. Former intersections have become grand plazas where young people play football and where there are table tennis and chess areas set up. Superblocks also created stretches of former streets that allow residents of certain areas like Barceloneta, the old fishing neighborhood, to have unimpeded access to beaches and the sea. Finally, according to the Ajuntament de Barcelona, since the inception of Superblocks, NO₂ and PM₁₀ particles have decreased by 25% and 17% respectively, in the area around the Sant Antoni Market. Similarly, noise pollution has decreased by 5% decibels in one of the most densely populated parts of the city. (Palència et al.) These considerable reductions in pollution substantially increase the quality of life for Superblock residents.

Before proclaiming Superblocks to be a grand success in accordance with Mediterranean cultures and values, it is important to examine who has benefited from their existence and who might have been harmed. For instance, only 16% of the population of Barcelona lives in the Eixample, the primary area where Superblocks have come into being. (Ajuntament de Barcelona [7]) By virtue of their design, Superblocks lend themselves best to densely populated areas where street layouts are regular and grid-like (Roberts [2]). In Barcelona, that specifically means the Eixample area in the very center of the city. The Ciutat Vella or ‘Old Town’ of Barcelona, while having high population levels, has mostly narrow and winding streets that create their own form of natural traffic calming. Farther away from the Eixample, but still in central Barcelona, areas like San Gervasi and Bonanova lack the particular density of population that helps to drive a successful Superblock model. And for residents of Barcelona in newer neighborhoods like Pedralbes or Les Corts, rolling out Superblocks has not even been contemplated due to their lack of population density.

It seems clear that those who live and own property in the newly enhanced neighborhoods of the Eixample are the principal beneficiaries of Superblocks. Large parts of Barcelona are noisy due to the din of cars and motorcycles zipping through narrow, canyon-like streets. The value of living in an area with less traffic and less noise is understandably appealing. Many residents are willing to pay a premium for the enhanced quality of life available in quiet, yet central, parts of the city. Those citizens who own property inside of the Superblocks have seen the financial value of their properties increase just by virtue of the Ajuntament deciding that their street was a prime location for gentrification. For example, on the prime central Superblock street Consell de Cent in 2023, price per square meter for properties is now 15% higher than in the non-Superblock Carrer de Valencia one block away (Casas) Others have been less fortunate.

Considering that 40% of the population of Barcelona do not own the homes in which they live, the higher property prices within Superblocks have not benefited everyone equally. (Palomera and González) Rental costs within Superblocks have increased dramatically. Using the same analysis as above, the difference between rental prices on Consell de Cent versus Carrer de Valencia are now EUR 3800 vs EUR 1950, respectively. Under the burden of price increases, long-term residents fear that they will have to move to other parts of the city (Nieuwenhuijsen et al.). Housing rights activists fear that further rent rises and gentrification will push out even more residents and will disturb the fabric of life which characterizes the Eixample. There are other beneficiaries, however, including those with no cars who prefer walking and cycling. Everyone who lives in the area and breathes cleaner air has also gained a benefit just from their homes being situated in the midst of a Superblock.

The issue of gentrification and its effects on residents of the Eixample is an important topic in Barcelona. The majority of Barcelona's citizens identify and vote as socialists. (Ajuntament de Barcelona [8]) Human rights and access to the basic necessities of life (like housing) are fundamental in the beliefs of citizens of this Mediterranean city. Before the creation of Superblocks, there was already a focus on locals being priced out of their living situations in the center as Barcelona (and many Mediterranean cities) became a popular tourist destination. Real estate speculators had been buying up buildings and apartments in the city center for years and converting them to Airbnbs and other forms of short-term lodging for the significant growth in tourism (Prinz). Leaders in Barcelona have countered by increasing investment in social housing for working- and middle- class residents, but as the city has only 1.5% of housing in these types of arrangements, it seems evident that there is a shortfall of public assistance available. (Ajuntament de Barcelona [9]) .

During early debates about Superblocks, residents and activists voiced their concern that those living on the outside of these new areas would suffer increased traffic jams along with more noise and pollution. It was also postulated that those who had cars and were living inside a new Superblock would experience difficulties in accessing and moving their vehicles. Barcelona's initial goal was to turn over 70% of the streets in the Eixample to pedestrians. Logically, then, the remaining 30% of open streets would have to carry the same number of cars. As data began to be understood, however, some of these concerns abated. The early Poblenou Superblock only saw an additional 3% traffic in the perimeter of its space. (Roberts [3]). It may be worth considering, however, whether there would be a compounding effect if the rest of Barcelona does actually become filled with green spaces created by Superblocks. The real savings in traffic, however, are the gains made in making walking and using public transportation more easily usable. This issue will not go quietly, however, as up to 60% of vehicular traffic is thought to come from commuters driving from regional areas into the city center.

Superblocks, ultimately, may not be an egalitarian solution to Barcelona's urban planning needs. With a broader perspective, however, it is clear that these creations came about through careful study, thought, and planning and were enacted by a government elected by the citizens of Barcelona. The intentionality around doing good for the majority of citizens could be seen as

evidence of a holistic embrace of the values that are shared and important in Mediterranean cultures. Threaded throughout each block, each park, each planted tree are the notions of what could be best for the majority of families, communities, and residents who live and enjoy these transformative urban spaces.

In sum, Superblocks may be viewed as being consistent with the fundamental tenets of Mediterranean life and culture, in that they have enhanced these values and ways of life. It can be difficult to unravel the composition of life in these areas into their fundamental components - importance of family, community, shopping, transport, and healthy living - but when it is done, a case can be argued for how Superblocks support each of these pillars. In examining a way of life that has persisted for two millennia, it seems clear that all of these fundamental elements have come to work in harmony with one another and with an integrated way of living for residents of Mediterranean cities. With projections that 68% of the world's population will be living in cities by 2050 (United Nations), it is interesting to think that Barcelona's Superblocks are just the type of modifications that are both consistent with Mediterranean culture and technologically adroit enough to propel Mediterranean city life ahead through the next millennium.

Works Cited

- Ajuntament de Barcelona [1] “Air Quality - Effects on Health.” www.Barcelona.cat, Ajuntament de Barcelona, 21 Nov. 2023, ajuntament.barcelona.cat/qualitativa/en/health/effects-health#:~:text=According%20to%20the%20studies%20carried,who%20live%20in%20the%20city.
- Ajuntament de Barcelona [2] “Urban Mobility Plan (2013-2018).” Urban Planning, Ecological Transition, Urban Services and Housing, Ajuntament de Barcelona, ajuntament.barcelona.cat/ecologiaurbana/en/what-we-do-and-why/active-and-sustainable-mobility/urban-mobility-plan. Accessed 29 Aug. 2024.
- Ajuntament de Barcelona [3] “The New Squares and Green Axes in Eixample Will Look Like This.” Superilles, Ajuntament de Barcelona, 14 Oct. 2022, ajuntament.barcelona.cat/superilles/en/content/the-new-green-hubs-and-squares.
- Ajuntament de Barcelona [4] “Fiestas y Tradiciones Por Distrito: Cultura Popular.” www.Barcelona.cat, Ajuntament de Barcelona, www.barcelona.cat/culturapopular/es/fiestas-y-tradiciones/fiestas-y-tradiciones-por-distrito. Accessed 29 Aug. 2024.
- Ajuntament de Barcelona [5] “Bike Lane Network.” Mobility and Transport, Ajuntament de Barcelona, 20 Nov. 2023, www.barcelona.cat/mobilitat/en/type-route.
- Ajuntament de Barcelona [6] “Green Hubs and Squares in the Eixample District.” Barcelona Superblock, Ajuntament de Barcelona, 18 Apr. 2024, www.barcelona.cat/pla-superilla-barcelona/en/green-hubs-and-squares-eixample-district.
- Ajuntament de Barcelona [7] “Edad Por Grandes Grupos de Edad: 1 Enero 2023.” Estadística y Difusión de Datos, Ajuntament de Barcelona, ajuntament.barcelona.cat/estadistica/castella/Estadistiques_per_temes/Poblacio_i_demografia/Poblacio/Padro_municipal_habitants/a2023/edat/edatg02.htm. Accessed 29 Aug. 2024.
- Ajuntament de Barcelona [8] “Election Results.” Arxiu Municipal de Barcelona, Ajuntament de Barcelona, 27 Aug. 2024, ajuntament.barcelona.cat/arxiuunicipal/en/virtual-exhibitions/exhibition-40-years-democratic-city-councils/election-results.
- Ajuntament de Barcelona [9] “Public Housing Stock.” www.Barcelona.cat, Ajuntament de Barcelona, 6 Apr. 2023, www.habitatge.barcelona/en/access-to-housing/public-housing-stock.
- “Barcelona, the City of Markets.” Meet Barcelona, Ajuntament de Barcelona, www.meet.barcelona/en/visit-and-love-it/barcelonas-markets#:~:text=Barcelona's%20network%20of%20municipal%20markets,four%20special%20non%2Dfood%20markets. Accessed 29 Aug. 2024.
- Blanchar, Clara. “Los Barceloneses Hacen La Mayoría de Las Compras En Su Barrio.” *El País*, 6 Apr. 2018, elpais.com/ccaa/2018/04/06/catalunya/1523009593_844012.html.
- Casas, Núria. “Speculation in the Eixample? Barcelona’s Superblocks Cause Housing Prices to Shoot Up.” *El Nacional*, 3 Oct. 2023, www.elnacional.cat/en/business/speculation-eixample-superblocks-housing-prices-barcelona_1103565_102.html.
- de la Torre-Moral, Andrea, et al. “Family Meals, Conviviality, and the Mediterranean Diet among Families with Adolescents.” *International Journal of Environmental Research and*

- Public Health, U.S. National Library of Medicine, 3 Mar. 2021,
www.ncbi.nlm.nih.gov/pmc/articles/PMC7967627/#B9-ijerph-18-02499.
- “El Municipio En Cifras.” Instituto de Estadística de Cataluña, Generalitat de Catalunya,
www.idescat.cat/emex/?id=080193&lang=es. Accessed 29 Aug. 2024.
- Laudani, Chiara, et al. “Does a Mediterranean Model of Family Functioning in the Perception of Italian and Spanish Adolescents Exist? A Cross-National Study.” *Mediterranean Journal of Social Sciences*, MCSER Publishing, Rome-Italy, Jan. 2014,
www.richtmann.org/journal/index.php/mjss/article/view/1915/1914.
- Matilla, Xavier. Personal interview. 21 May 2024
- Muñoz, Elena. “Pasear Como Método de Desconexión Para Aliviar La Carga Mental Del Día a Día.” *El País*, 22 Mar. 2024,
elpais.com/estilo-de-vida/2024-03-20/pasear-como-metodo-de-desconexion-para-aliviar-la-carga-mental-de-l-dia-a-dia.html.
- Nieuwenhuijsen, Mark et al. “The Superblock Model: A Review of an Innovative Urban Model for Sustainability, Liveability, Health and Well-Being.” *Environmental Research*, Academic Press, 1 Mar. 2024,
<https://www.sciencedirect.com/science/article/pii/S0013935124004547?via%3Dihub>
- “One out of Every Two Inhabitants of Barcelona Already Uses Smou.” *SMOU*, 28 Feb. 2023,
www.smou.cat/en/news/4th-anniversary-smou.
- Palomera, Jaime, and Jordi González. “Myths and Truths about Generation Rent: Barcelona Metropolis: Barcelona City Council.” *Barcelona Metropolis*, 9 Oct. 2023,
<https://www.barcelona.cat/metropolis/en/contents/myths-and-truths-about-generation-rent#>
- Palència, Laia, et al. “Report of Results of the Environmental and Health Effects of the Superblocks Model in Barcelona.” *Salut Als Carrers Project*, Agència de Salut Pública de Barcelona, 2021,
www.aspb.cat/wp-content/uploads/2021/10/English-ASPB_salut-carrers-resultsreport-Superblocks.pdf.
- Prinz, Ulrike. “Barcelona’s Residents Want Their City Back.” *Goethe-Institut Madrid*, May 2018, www.goethe.de/ins/gb/en/kul/ser/past/liv/21275061.html.
- Planella, Ariadna Baró. “Barcelona Gives Way to Green Infrastructure.” *diVA Portal*, 2019,
www.diva-portal.org/smash/get/diva2:1333820/FULLTEXT01.pdf.
- Roberts, David [1]. “SUPERBLOCKS - BARCELONA’S PLAN TO FREE ITSELF FROM CARS .” *Kleinman Center for Energy Policy*, June 2019,
kleinmanenergy.upenn.edu/wp-content/uploads/2020/08/Superblocks-1.pdf.
- Roberts, David [2]. “Could Barcelona’s Plan to Push out Cars and Build Superblocks Work in the US?” *Vox*, 12 Apr. 2019,
www.vox.com/energy-and-environment/2019/4/12/18273897/barcelona-urban-planning-portland-oregon-superblocks.
- Roberts, David [3]. “Barcelona Wants to Build 500 Superblocks. Here’s What It Learned from the First Ones.” *Vox*, 9 Apr. 2019,
www.vox.com/energy-and-environment/2019/4/9/18273894/barcelona-urban-planning-superblocks-poblenou.
- Rueda, Salvador. Personal interview. 25 June 2024

- “Spain Is the World’s Healthiest Nation, According to Bloomberg Index.” EL PAÍS , 26 Feb. 2019, english.elpais.com/elpais/2019/02/25/inenglish/1551113690_247893.html.
- “Story of Cities #13: Barcelona’s Unloved Planner Invents Science of ‘Urbanisation.’” The Guardian, Guardian News and Media, 1 Apr. 2016, www.theguardian.com/cities/2016/apr/01/story-cities-13-example-barcelona-ildefons-cerda-planner-urbanisation.
- Tatjer, Mercedes. “LA VIVIENDA POPULAR EN EL ENSANCHE DE BARCELONA.” Scripta Nova: REVISTA ELECTRÓNICA DE GEOGRAFÍA Y CIENCIAS SOCIALES, Universidad de Barcelona, 1 Aug. 2003, [www.ub.edu/geocrit/sn/sn-146\(021\).htm](http://www.ub.edu/geocrit/sn/sn-146(021).htm).
- TMB, Fundació. “Cronología.” TMB, Grupo TMB, historia.tmb.cat/100-anys-xarxa-bus/es/cronologia/. Accessed 29 Aug. 2024.
- “The Birth of a Nation. from the 8th to the 13th Century.” mhcat.cat, Museu d’Història de Catalunya, www.mhcat.cat/enmhc/exhibitions/the_memory_of_a_country/the_birth_of_a_nation_from_the_8th_to_the_13th_century. Accessed 29 Aug. 2024.
- United Nations, “68% of the World Population Projected to Live in Urban Areas by 2050, Says UN.” United Nations Department of Economic and Social Affairs, United Nations, 16 May 2018, www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.
- Urbano, Judith. “The Cerdà Plan for the Expansion of Barcelona.” The Cerdà Plan for the Expansion of Barcelona: A Model for Modern City Planning, digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1323&context=focus.
- Wynn, Martin. “Barcelona: Planning and Change 1854-1977.” The Town Planning Review, vol. 50, no. 2, 1979, pp. 185–203. JSTOR, <http://www.jstor.org/stable/40103366>.

Can Super High School Construction Promote Economic Growth? Panel Data Based on Provincial Level By Xiaobo Zhou

Abstract:

Educational equality is an essential foundation of social equity, but the problem of educational inequality in China is still very prominent. The emergence of super middle schools, centralization of education, and monopolization of educational resources affects the level of human capital accumulation in society, subsequently impacting economic growth. Super high school is a special education phenomenon produced by a specific stage of educational development in China in 1999. However, with the rapid economic development, people are increasingly eager for educational resources. The crave for resources results in unfair and unjust phenomena in the social development of “super schools,” sparking criticism and doubts from the public.

Based on the provincial panel data of 31 provinces, autonomous regions and municipalities in China from 2012 to 2020, this article measures the monopoly rate of super high schools in various provinces and analyzes the current status of education centralization and income inequality. Through establishing a two-way fixed effects model, I will show how educational inequality perpetuates intergenerational poverty. The analysis reveals that when the concentration of education is lower than a certain value, the concentration will have a positive impact on the economy. However, when it is excessive, it will become unfavorable to economic development.

Keywords Education monopoly, Super high school, educational centralization, economic growth, elite universities

Introduction

China attaches great importance to the accumulation of educational human capital. A high level of educational human capital means that people have a high level of education, and therefore a higher mastery of skills, work efficiency and production quality. Many studies also show that educational human capital is a long-term driving force to support sustained and stable economic growth. In 1986, China decided to implement nine-year compulsory education. In 1995, the state again put forward the strategy of "rejuvenating the country through science and education", which emphasized that education and science and technology were important factors to promote economic development. The continuous reform of the education system has expanded the scale of higher education and promoted the rapid development of China's economy. However, since the reform and opening up, the income gap of Chinese residents has continued to expand (Li and Zhu 2018; Xue and He 2015). The intergenerational cycle of poverty appears (Han 2010; Huang 2014; Lin and Zhang 2012), and the social stratum tends to solidify (Zhang 2012; Yang 2015).

In this modern society, with the continuous expansion of the scale of higher education, the academic qualifications have become increasingly devalued. In the list of outstanding alumni of Chinese universities in 2020, Peking University and Tsinghua University, the top two universities in China, produce the highest number of outstanding talents in the political and business fields in China. The average salaries of their graduates are also far ahead of those of graduates from other schools. At the same time, with high-quality students, teachers, abundant education funds and good infrastructures, super high schools increasingly monopolize the enrollment opportunities of elite universities. Because education has strong intergenerational transmission and can solidify social production relations, the educational gap also tends to be transmitted to the next generation. The centralization and monopoly of education, along with inequalities in educational opportunities, have led to a weakening of intergenerational mobility and a solidification of educational and social class structures. This has resulted in the current situation where individuals at the bottom of Chinese society are abandoning education, the middle-class experiences excessive anxiety, and the upper class expresses dissatisfaction with the education system (Guo and Xia 2021). The existence of problems such as the high cost of obtaining higher education and the obstruction of the poor class in trying to change their social status has had an inestimable negative impact on economic development and social welfare in the long run.

Literature Review

2.1 The Significance of Education

The United Nations Education Scientific and Cultural Organization advocates the importance of humanism in the context of globalization. It reflects on the kind of people education should cultivate, and points out that education should respect life, personality, peace, equality, human rights and interests, and educational institutions should take responsibility for sustainable development. UN's advocacy helps us rethink education and its importance from the social, economic, and environmental aspects. It highlights that education should fulfill its own role and adapt to globalization. Wang Mo and others (2012) repeatedly emphasized humanism as the core of education in their research in reflective education, and advocated that under the guidance of humanism, education should be the universal solution to the problems arising under the globalization situation. Tang and Huang (2020) expound the humanitarianism and humanitarianism concepts of the United Nations Education Scientific and Cultural Organization. They discussed the rationality of humanism from the three levels—individual, educational, and social—and emphasized the importance of the humanistic education concept in the backdrop of the change of educational patterns worldwide. They also emphasized the importance of the humanistic education concept from the perspective of enhancing educational inclusiveness. This paper illustrates the concrete path of practicing the humanistic education concept proposed by “reflective education” from three perspectives, such as a comprehensive education method.

2.2 the Impact of Education Level on Economic Growth

As for the impact of education level on economic growth, both theories and practices arrive at a relatively consistent conclusion. That is, education level is positively related to economic growth. Moreover, the early human capital theory shows that education can lead to higher income. The research of Ding and Knight (2011) shows that the increase in the enrollment rate of middle schools and colleges after China's reform and "Opening-up" has greatly promoted China's economic growth. Therefore, expanding the scale of education or increasing the public education expenditure can reduce income inequality and promote economic growth. However, according to empirical evidence, expanding the scale of education has widened the income gap, because the career prospects of those who attended higher education are far better than those who stopped at compulsory education.

Scholars have done a lot of research on the inequality of educational opportunities. Many focus on the relationship between income inequality and education inequality. They argue that education inequality contributes to the widening income distribution gap, and education has become an internal mechanism to solidify and perpetuate income inequality (Wen 2011), which ultimately results in serious class solidification. Li (2010) analyzed the hidden inequality within higher education after the expansion of university enrollment, indicating that whether children can receive higher education is closely related to their family background. Education inequality can have a sustained impact on the urban-rural income gap through intergenerational transmission (Zhang 2006). In short, education expansion cannot reduce income inequality (Mincer 1974).

Income inequality hinders technological progress and affects human capital accumulation, thus hindering the improvement of economic development. In terms of demand, income inequality reduces the proportion of demand in consumption in GDP, thus hindering economic growth. However, it also increases the proportion of total capital and total net exports in GDP, thus promoting economic growth. Liu Haiying (2004) and Chen Zhongchang (2011) studied the cross-sectional data of China's population census and found that the Gini coefficient of the regional education level is negatively related to the average education level of the region. Therefore, promoting equal distribution of education levels is conducive to improving the overall education level, leading to a growth in economics. However, the uneven distribution of education is not conducive to the development of education. With the increase of the population participating in higher education, the technological level and innovation level of the region will also be improved, thus promoting economic development. When the education level of the overall labor force in the region is low, it will lead to a low level of technology and a decline of industrial innovation ability in the region, which will ultimately hinder the improvement of the regional economic development level.

2.3 The Super High school

China has entered the stage of popularization of higher education. Only when rural students enter elite universities can they realize the advancement of social status and block

intergenerational poverty. However, nowadays, the opportunities for rural students to enter elite universities are decreasing. Influenced by the structural expansion of higher education and other factors, in recent years, super high schools are increasingly monopolized by virtue of high-quality students, teachers, abundant education funds, good infrastructures, and the opportunity for admission to elite universities. The first result is that inter school education is becoming more and more unfair. That is, the emergence of super high schools has caused the decline of the quality of education in other schools, especially in the county high schools. Secondly, super high schools also contribute to class solidification. For example, students from high-income families are more likely to obtain high-quality high school education resources. The third result lies in the unequal treatment of students in super high schools. In order to pursue the enrollment rate, super high schools may excessively favor top students in the allocation of school education resources (Yang 2015; Guo and Wang, 2018; Gan and Liu, 2015; Li and Cui, 2016).

2.4 Privatization of Education

On the other hand, with the introduction of commercial operation mode into public education, private education tends to be profit-making. The nature of education has changed from a public welfare undertaking to a private or consumer goods (United Nations Education Scientific and Cultural Organization). At present, the marketization trend has made the boundary between public and private education more blurred. The privatization of education has also led to the increasingly serious problem of education centralization and the increasingly serious class solidification. Behind the lack of efficiency and fairness in China's higher education is a monopoly and interest dispute in Higher Education. Eliminating the structural obstacles to the development of higher education and realizing the efficiency and freedom of the higher education market are crucial for ensuring the sustainable development of China's economy.

Research Design

Data

This essay adopts the panel data from 31 Chinese provinces from 2012 to 2020 so that it forms balanced data to indicate the relationship between monopoly rate of super high school and economic development. The data stems from the China Statistical Yearbook. The data source of educational monopoly rate originates from another article (Guo and Xia 2021), which estimates the educational monopoly rate of super high schools in 31 provinces.

Measurement

Dependent Variable

This essay implements GDP per capita as the dependent variable. Choosing GDP per capita as a measure of economic growth is the best choice based on data availability and other aspects (Lu Ming, 2020). GDP per capita is a set of national economic accounting methods with

market value as its core. The logic of this method is rigorous, and a complete system is established with quantitative indicators. It has established a basic framework for digital management of the social economy. Obviously, the measurement results are more accurate and objective. In this sense, GDP per capita is the most important and comprehensive economic indicator we can see at present. In order to make economic growth more realistic and eliminate the influence of price factors, this paper takes the natural logarithm of GDP per capita.

Independent Variable

The monopoly rate of super high schools is an independent variable. It is able to be calculated by the following formula:

$$super_{it} = \frac{S_{it}}{N_{it}}$$

Among them, S_{it} is the number of super high school students admitted by University A in the T year of the I Province, N_{it} is the total number of high school students admitted by University A in the T year of the I Province, and $super_{it}$ is the proportion of super high schools in each province in the enrollment quota of University A in that year, and the value is between 0 and 1. The lower the value, the lower the monopoly degree of super high schools on the enrollment opportunities of elite universities in the province, and other ordinary high schools have more enrollment opportunities of elite universities; A higher the value means that the enrollment opportunities of elite universities were excessively concentrated in a few super high schools in the province, and the enrollment opportunities of other ordinary high schools were largely occupied (Guo and Xia 2021).

Control Variable

Typically, I control three variables that affect the development of the economy: population, industrial structure, and globalization level. According to previous studies, population is the key factor influencing economic development. The relationship between population and economic development is very close. As we all know, regional economic development will promote "population aggregation", and in turn, "population aggregation" brings "aggregation effect" to the economy. Population distribution is the form of long-term economic and social development, which can be reflected in population size, population quality and population structure (He 2022). Second, the level of economic development is closely related to the industrial structure. In today's complex and changeable domestic and international situation, it is an important task to stabilize China's economy by building a modern economic system, accelerating the upgrading of industrial structure, and breaking through the supply constraints (Zhu 2022). Additionally, globalization has become an

unavoidable economic trend and is an important driving force to support the economic growth and national economic development of all countries (Han 2022).

Model

It is uncertain whether and how such factors affect economic production. This study adopts two-way fixed effect model to examine the impact of monopoly of super high school on GDP per capita the following model:

$$\text{Ln}Y_{it} = \beta_0 + \beta_1 \text{LnSchool}_{it} + \beta_2 \text{LnSecond}_{it} + \beta_3 \text{LnUrban}_{it} + \beta_4 \text{LnRoad}_{it} + \mu_i + \varepsilon_{it}$$

Y_{it} represents the economic strength, namely GDP per capita; School_{it} refers to the monopoly of super high school; Second_{it} represents the amount of second industry production of the public; Urban_{it} represents the urbanization rate of the province; Road_{it} refers to the highway mileage in each province. μ_i refers to the fixed effect of individual provinces; ε_{it} refers to the disturbance item.

Results

Descriptive Analysis

Table 1 shows the descriptive analysis of our dependent variable, independent variables, and control variables. Additionally, there are also the mean, standard deviation, minimum and maximum values corresponding to each data. Firstly, the dependent variable is GDP per capita, the mean of counterpart is 30873 yuan, and the standard deviation is 20354, which indicates that the regional economic developments of various provinces are drastically different. The maximum difference has reached 10094 yuan. The mean value of the monopoly rate of super high school is 0.29, and the standard deviation is 0.16. The centralization of education is most obvious in Qingdao, which has a maximum value of 0.89. In contrast, Shandong contains the lowest monopoly rate of super high schools, only 0.02. In addition, Xinjiang, Hainan, and Guizhou also have relatively high educational monopoly rates.

Table 1 Summary table

Variable	Definition	Mean	Std. Dev.	Min	Max
GDP	GDP per capita	30873.50	20353.77	4317.00	105231.00
School	monopoly rate of super high sch	0.29	0.16	0.02	0.89
Second	Second industry production	5811.27	5961.04	51.30	31930.40
Urban	Urbanization rate	0.50	0.15	0.21	0.90
Cap	Financial capacity	0.50	0.21	0.06	0.95
Death	Population mortality	5.95	0.68	4.21	7.28
Old	Old age dependency ratio	0.12	0.02	0.07	0.20
Road	Highway mileage	11.61	6.94	0.78	30.97

Regression analysis

In model 1, there are three variables—second industry production, urbanization rate and highway mileage. Among them, the urbanization rate has a significantly positive impact on the dependent variable, which means the higher the level of urbanization, the more GDP per capita. Additionally, the highway mileage also had a positive effect on GDP per capita. However, the influence is much lower than the former. The second industry production has almost no impact. In model 2, we added financial capacity and old age dependency ratio. We found that the financial capacity has a positive effect but not significant and the old age dependency ratio gives a remarkable positive effect. This means when the two are improved, the GDP per capita will increase. Last but not least, we added the key independent variable into model 3, and it can affect the GDP per capita negatively. In other words, when the monopoly rate of super high schools increased, the development of the economy was hindered.

Table 2 Regression analysis of the impact of monopoly of super high school on Economic development

	(1)	(2)	(3)
monopoly rate of super high school			-0.127** (0.053)
Second industry production	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
Urbanization rate	1.119*** (0.192)	0.940*** (0.181)	0.950*** (0.179)
Ln (Highway mileage)	0.221*** (0.038)	0.187*** (0.035)	0.198*** (0.035)
Financial capacity		0.491*** (0.142)	0.477*** (0.141)
Population mortality		0.015 (0.015)	0.014 (0.015)
Old age dependency ratio		2.784*** (0.422)	2.763*** (0.419)
Constant	8.546*** (0.091)	8.028*** (0.139)	8.054*** (0.138)
Year	Yes	Yes	Yes
Observations	341	341	341
R-squared	0.975	0.979	0.980
Number of pro	31	31	31

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Discussion and Conclusion

Education serves as the bedrock of national strength, while the economy underpins development. The interplay between education and economic growth is crucial, with each influencing and supporting the other. As market economies evolve, education must adapt to avoid exacerbating social inequalities like class stratification. Super high schools in China exemplify the tensions between educational centralization and market forces, highlighting the need for reforms to mitigate class solidification. This study utilizes provincial panel data to analyze the relationships between monopoly rates of super high schools, per capita GDP, and financial capacities, revealing significant regional variations in education centralization and super high schools' impact on economic growth.

Addressing the issues posed by super high schools requires us to prioritize the equitable distribution of educational resources. Balancing resources across high schools and phasing out super high schools and specialized institutions are essential steps. The recent revisions to the compulsory education law aim to promote balanced development and equalize school conditions, moving away from distinguishing between key and non-key schools. A robust education system should ensure equal access to quality education for all students, fostering class mobility and reducing inequality.

Works Cited

- Baogui, Xue, and He Liancheng. "A Review of Research on Income Inequality of Chinese Residents." *Economist*, vol. 5, no. 2, 2015, pp. 82-90.
- Chun, Han. "Research on the Root Cause of Intergenerational Transmission of Rural Poverty in China." *Economic Research Guide*, no. 16, 2010, pp. 46-48.
- Congbin, Guo, and Wang Jiaqi. "Where is the Source of Students in China's Elite Universities – Taking the 2013 Students of A University and B University as an Example." *Education Research*, no. 12, 2018, pp. 99-108.
- Congbin, Guo, and Xia Yufeng. "The Impact of Super Middle Schools on the Enrollment Opportunities of Rural Students in Elite Universities." *Higher Education Research*, no. 7, 2021, pp. 45-57.
- Dandan, Guo. "The Genetic Mechanism and Transformation Logic of Educational Inequality." *Journal of East China Normal University (Educational Sciences)*, vol. 40, no. 6, 2022, pp. 116-26.
- Ding Sai, and John Knight. "Why has China Grown So Fast? The Role of Physical and Human Capital Formation." *Oxford Bulletin of Economics and Statistics*, vol. 73, no. 2, 2011, pp. 141-74.
- Dongping, Yang. "'Super Middle School' Goes Beyond the Law of Education." *China Daily*, 23 January 2015, china.chinadaily.com.cn/shizheng/2015-01/23/content_19390087.htm. Accessed 30 August 2024.
- Haifeng, Zhang. "Urban and Rural Education Inequality and Widening Income Gap: An Empirical Analysis Based on Provincial Mixed Cross-sectional Data." *Journal of Shanxi University of Finance and Economics*, no. 2, 2006, pp. 31-38.
- Jiaoxiu, Wen, and Wang Yanjun. "A Dynamic Study on the Widening Gap between Education Inequality and Income Distribution in China – An Empirical Study Based on the Gini Coefficient of Education in Various Regions." *Journal of Chengdu University of Technology (Social Science)*, no. 1, 2011, pp. 5-10.
- Jie, He. "Research on the Impact of Population Distribution on High-Quality Economic Development." *Statistics and Consultation*, no. 3, 2022, pp. 27-31.
- Jingxiao, Huang, and Tang Xiaomeng. "Modernization and Implications of Japanese Higher Education towards 2040." *Journal of Modern Education*, no. 6, 2020, pp. 72-77.
- Junren, Liu et al. "Difficulties and Breakthroughs in Cultivating Full-Time Master of Education in Endogenous Comprehensive Universities." *Journal of Henan Institute of Education (Philosophy and Social Science)*, no. 4, 2019, pp. 46-53.
- Lidong, Zhang. "Empirical Study on the Intergenerational Transmission of Poverty in Rural China." *China Population Resources and Environment*, vol. 23, no. 154(6), 2013, pp. 45-50.
- Mingang, Lin, and Zhang Ruili. "Research on Intergenerational Transmission of Rural Poor Families – Analysis Based on CHNS data." *Agricultural Technology and Economy*, no. 1, 2012, pp. 29-35.

- Qunmei, Zhang. "The Analysis of Class Mobility Solidification Trend and Governance Path in Current Chinese: Group Interest Perspective." *Journal of Henan University (Social Science)*, vol. 52, no. 246(3), 2012, pp. 34-39.
- Shi, Li, and Zhu Mengbing. "Changes in the Income Gap of Residents in the 40 Years of China's Economic Transformation." *Management World*, no. 12, 2018, pp. 19-28.
- Wenwei. Yang. "The Negative Effect of Class Solidification and Its Way Out." *Journal of the Party School of the Central Committee of the Communist Party of China*, no. 1, 2015, pp. 101-6.
- Xiao, Huang. "How to Prevent the Matthew Effect of Poverty – The Perspective of Intergenerational Income Flow." *Economic Management*, no. 5, 2014, pp. 15.
- Xiaomeng, Tang, and Huang Jingxiao. "The Reaffirmation of the Humanistic View of Education – Interpretation of the UNESCO Report on Reflective Education." *Higher Education Exploration*, no. 8, 2017, pp. 49-53.
- Xingdong, Li, and Cui Mengtian. "Analysis of the Phenomenon of 'Super Middle School' from the Perspective of Sociology." *Educational Science*, no. 5, 2016, pp. 26-30.
- Yafu, Tu. *Education Inequality and Its Impact on Economic Growth*. 2016. Nanjing Audit University, Master's thesis. CNKI, kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201801&filename=1018021833.nh.
- Yang, Han. *Research on the Development Problems and Countermeasures of "Super Middle School"*. 2022. Hebei Normal University. Master's thesis. CNKI, kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFDTEMP&filename=1022487814.nh
- Ying, Zhu. "Mechanism Analysis of Industrial Structure Adjustment in the Context of Economic Transformation." *Industrial Innovation Research*, no. 14, 2022, pp. 102-4.
- Zhiwen, Han. "Research on the Relationship between Economic Globalization and Income Distribution – Based on the KOF Economic Globalization Index." *China Prices*, no. 9, 2019, pp. 29-31.

Estimating Bulgaria's 2023 Bright Economy Composite Index: Insights Into the Size of Hidden Economy By Aleksandar Georgiev

Abstract

This study estimates Bulgaria's Bright Economy Composite Index (BECI) value, a crucial tool for assessing the country's transparency and formalization of economic activities. By analyzing both the statistical and sociological indicators that BECI comprises, the paper examines the trends in Bulgaria's bright economy and infers about the scope and size of the country's hidden economy. This research paper employs four methods of statistical estimation in order to predict BECI for 2023 based on past data and trends, as well as multiple scenarios that could have caused the composite index to deviate from its natural progression. By utilizing different variations and extracts from the complete dataset to ensure robustness and accuracy, this study forecasts that the estimated BECI 2023 value will be in the 78.19-80.50 range, implying the hidden economy accounts for about 20% of economic activity. The paper's findings suggest that while Bulgaria has made progress in brightening its economy, significant challenges remain, particularly in sectors prone to informal activities, highlighting the need for continued governmental efforts to reduce the hidden economy through targeted policies and increased transparency. The paper concludes by outlining potential avenues for future research, such as expanding the application of BECI to other sectors, exploring cross-country comparisons, and examining the impact of digital transformation on the hidden economy.

1. Introduction

Economies have conventionally been divided into the "bright" and the "hidden" economy. The bright economy comprises all activities that are officially declared and which, therefore, contribute to forming a part of the GDP. On the other hand, the hidden economy comprises undeclared activities ranging from home production and illegal businesses to legal activities concealed from authorities. A more significant share of GDP by the bright sector suggests that the economy is less dependent on illegal or government-concealed activities and, therefore, creates a more transparent and law-abiding economic setting that fosters trust and economic stabilization. In contrast, a higher value of the proportion of the hidden economy could indicate higher levels of tax evasion, informal labor, and unreported economic transactions. Monitoring the distribution of GDP between the bright and the hidden sectors is significant to countries as it affects policymaking and helps derive inferences about economic growth. More accurate data on the hidden economy can contribute to designing strategies for shifting activities from the hidden to the bright economic sectors, also known as "brightening" the economy. Improved transparency will likely result in more tax revenues and increase overall economic stability. In this light, national policymakers can design interventions to encourage formalization and ensure sustainable development. In this paper, I analyze the case of Bulgaria, which has been measuring the size of its bright economy since 2010 using an index-based measure.

The Bright Economy Composite Index (BECI), created by the Bulgarian Industrial Capital Association (BICA), evaluates the dynamics of Bulgaria’s formal economy’s spread. The index was crafted to monitor the bright economy’s patterns and trends, and its annual values help economists infer relative economic growth, expansions, or shrinkages in this sector of the economy. The foundations of BECI lay upon the perception that certain economic activities occur “below the surface” and are unreported by government agencies, and so allow official and informal (unreported or “hidden”) economies to coexist [1]. In its essence, the so-called hidden economy includes economic activity that is not reported in official statistics; the broadest definition of the hidden economy encompasses home production, illegal economic activities, and legal economic activity that is purposefully kept hidden from measurement, which has consistently been making assessing the actual size of this economy a challenging task [2].

While every modern economy has an “underground” or “hidden” sector, this phenomenon occurs to varying extents, and although challenging, evaluating the scope and dynamics of “informality” as an economic issue is crucial for developing sound economic policies. Throughout recent decades, numerous studies have placed Bulgaria among the EU countries with the most significant proportion of GDP sourced from the hidden sector. In 2022, according to *ResearchGate*, Bulgaria ranked highest in Europe, with an estimated 33.1% of the overall annual GDP sourced from activities within the said sector, almost double the European average of 17.3% [3].

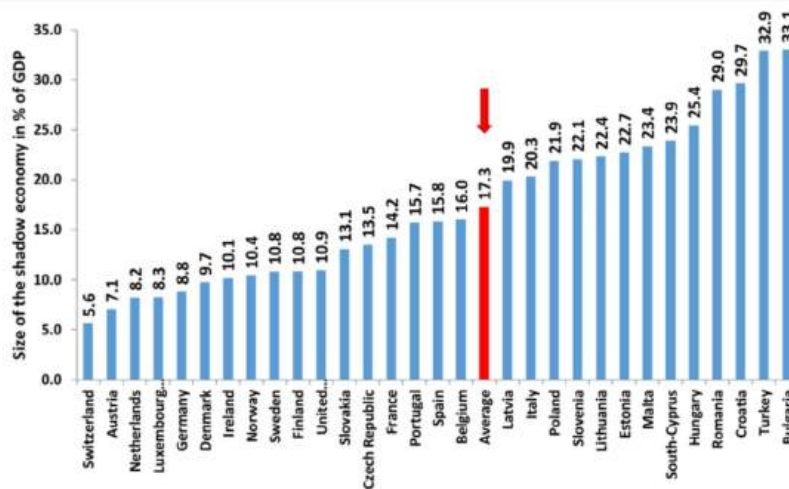


Fig. 1: Size of the hidden economy of 31 European countries in 2022 (in % of GDP) [3]

Over the years, the Bulgarian government has attempted to implement numerous policies that would incentivize people to “brighten” their business activities with the long-term aim of decreasing the relative size of the hidden economy. Since its creation, BECI has been essential in dealing with the high rate of Bulgaria’s shadow economy. BECI values range from 0-100%, making inferences about the hidden economy simple. Using BECI, in simple terms, informs the public about the potential size and consequences of the shadow economy and develops businesses through regulation, thereby rendering government involvement in policing more efficient. BECI has become a means through which formal pathways to policy work are opened,

ensuring enhanced transparency through rigorous information gathering and policy intervention awareness.

In calculating the final annual BECI value, two categories of indicators are considered: statistical and sociological. The statistical category comprises seven indicators, the raw data sourced indirectly from official statistical databases such as the National Statistical Institute and the Bulgarian Ministry of Finance. The sociological category of indicators branches into two subcategories, one related to employers (eleven indicators) and the other - to employees (seven indicators).

In this research paper, I employ four methods of statistical estimation in order to predict BECI for 2023 based on past data and trends, as well as consider multiple scenarios that could have caused the composite index to deviate from its natural progression. By utilizing different variations and extracts from the complete dataset to ensure robustness and accuracy, this study forecasts that the estimated BECI 2023 value will be in the 78.19-80.50 range, implying the hidden economy accounts for about 20% of economic activity.

2. Background

Established in 2010, the National Center for Bright Economy aims to decrease and avert all forms and implications of the informal economy, including undeclared employment, by implementing methodical measures for sustainable national development. Since 2014, The National Center for Bright Economy has been calculating annual BECI values using “the innovative EU-level tool to measure the bright side of the economy and the trends and challenges to mitigate and prevent the informal economy” [4].

The data attributed to the sociological category of BECI is sourced directly, primarily through surveying the two social groups that form the subcategories: employers and employees. According to a spokesperson from the Bulgarian Industrial Capital Association, a constant panel of the same respondents is used to source data annually. These panel respondents include representatives of at least 600 enterprises of all sizes and a minimum of 1280 employed individuals. All cumulative values are measured in percent [%]. Some of the indicators include measures that aim at estimating the extent to which businesses and individuals in Bulgaria adhere to formal labor contracts, receive full remuneration, and comply with tax regulations. Additionally, the indicators assess the frequency and accuracy of financial transactions, such as payments and VAT declarations, and the presence of formal systems, like quality control and proper receipt issuance for goods and services. These measures collectively offer a comprehensive overview of economic transparency, compliance, and the formalization of economic activities in the country.⁵

After the creation of BECI in 2014, using data sourced by the National Statistical Institute and the Ministry of Finance, scientists from BICA performed the calculations for every year from 2010 to 2014. The creation of the composite index took a rigorous three-year discussion, and the components were possibly incorporated by referencing other countries that

⁵ Complete list of indicators available in section 3.1

have been measuring the size of the hidden economy and the tax-evading policies they have issued.⁶ In assessing the size and dynamics of this sector, researchers have outlined each component with the intention of minimizing the GDP leakage margins that are not included in the calculations. For example, if not all cash payments are reported by an enterprise, there could potentially be a leakage in the GDP measure. Similar is the scenario where, if one does not report all the labor contracted, expenditure remains underreported, and so does the employee’s income, which will eventually remain outside the GDP measurement. The “Declared VAT” indicator, likewise, is taken into consideration because it assesses the declared invoices by the economy. Anecdotally, we know that some businesses may be generating fictitious invoices that never get reported in their sales, and therefore, the company never pays VAT to the government. A 100% value of declared VAT would mean that every economic sale is invoiced. However, surveys conducted in 2001 show that “almost 17% of corporate tax is evaded by purchasing fictitious invoices” [6]. In their annual reports on BECI and its implications, BICA outlines that each indicator’s measure is directly related to the GDP and could lead to a gap in the overall GDP measure if not considered, analyzing each indicator separately in a similar manner. In Bulgaria, BECI has aided in economic expansion and improved transparency. In simple terms, BECI assesses the condition of the national economy in conjunction with sector-specific transparency, making it easier to identify trends and shape policies that would boost economic stability. Holistically, BECI stimulates sales, increases sectoral efficiency, promotes transparency and sustainability, and helps policy-making through reliable research and investment ideas.

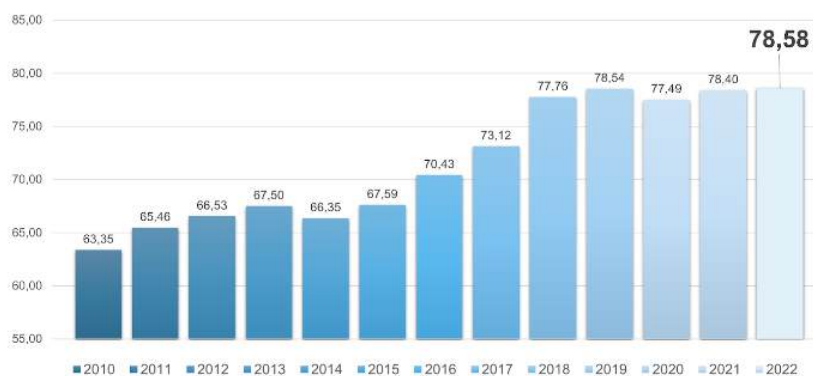


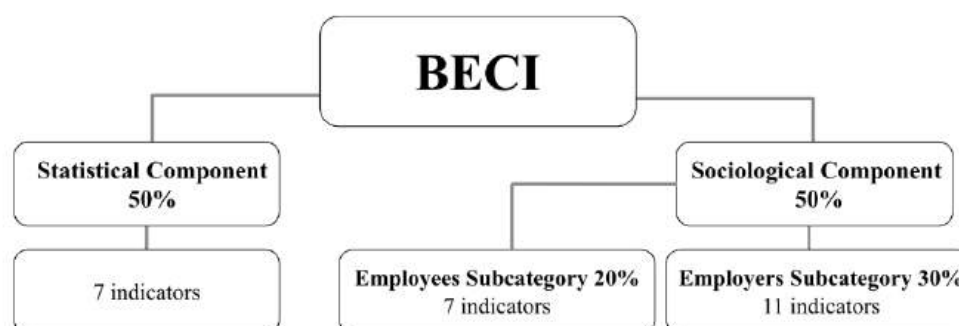
Fig. 2: BECI Dynamics [5]

The bar chart in Fig. 2 exhibits the growth of the BECI values since 2010. Since 2015, there has been a steady annual increase of about 3% up until 2018, when there was a larger increase of 4.6% in the BECI value compared to 2017. Since 2018, BECI values have been progressing relatively proportionally once again with a 0.1% margin, except for 2020, where the economic situation took a turn due to COVID-19 in Bulgaria and throughout the world.

⁶ Due to the limited information available on the historical factors behind BECI, this section primarily uses authorial assumptions based on articles officially published by BICA.

The latest officially published value of BECI is 78.58, measured in 2022. As the hidden economy is identified as the complement of the BECI value each year, it could be concluded that it accounts for $(100-78.58)\% = 21.42\%$ of Bulgaria’s GDP for 2022 [5]. Although Bulgaria has been demonstrating adequate progress in brightening the economy and decreasing the proportion of GDP coming from “underground” activities, Bulgaria has to continue investing efforts and implementing tailor-made policies to be evaluated as closely to the 17.3% European average as possible.

3. Methodology & Data



Flowchart 1: BECI Indicator Breakdown by Components

3.1 Indicator Meaning, Measurement, and Weight

The table below explains each indicator that is utilized in the calculation of the index. The statistical indicators primarily look into the overall nature of transactions in the economy. For example, we can see the proportion of declared VAT (Value Added Tax) has a 13.5% weight in the index. One could imagine that this measure indicates how close the reported VAT is to the actual expected VAT in the economy; a higher proportion would thus represent that the economy is “brighter.” In other words, it would mean that the extent of hidden transactions or unreported transactions on which VAT would be evaded is lower.

Description of each indicator, specifying that the units are proportions, in the 0-100 range, and explanation of the proportion they measure:

Statistical Components	Weight	Description
Cash Payments	20.00%	The proportion of total transactions made in cash out of all transactions.
Declared VAT	13.56%	The proportion of the total Value Added Tax (VAT) declared by businesses out of the expected VAT.
Socially Secured Employed	5.50%	The proportion of employees with social security coverage out of the total workforce.
Informally Employed	N/A	The proportion of workers engaged in economic activities without official contracts, social security registration, or legal recognition.

Hired By Contract	1.58%	The proportion of employees hired under formal contracts out of the total workforce.
Declared labor resource	5.50%	The proportion of the total labor force declared by businesses out of the total potential labor force.
Excise duty		
Excise duty on tobacco	3.28%	The proportion of excise tax revenue from tobacco products out of the total expected excise tax revenue from tobacco.
Excise Duty on alcohol	0.28%	The proportion of excise tax revenue from alcoholic beverages out of the total expected excise tax revenue from alcohol.
International Trade		
Trade within the EU	0.20%	The proportion of trade transactions within European Union countries out of the total possible trade transactions within the EU.
Trade outside EU	0.10%	The proportion of trade transactions outside European Union countries out of the total possible trade transactions outside the EU.
Subtotal:	50%	
Sociological Components	Weight	
Employers	30.00%	
Labor Market - labor with a contract	1.70%	The proportion of the labor force with formal employment contracts out of the total labor force.
Labor Market - full remuneration labor	1.70%	The proportion of the labor force receiving full remuneration out of the total labor force.
All Production Declared	3.30%	The proportion of the total production output declared by businesses out of the total expected production output.
Legal Import and Export	3.30%	The proportion of import and export activities that are legally compliant out of the total import and export activities.
Correct Duty and Tax Declaration	3.30%	The proportion of businesses accurately declaring duties and taxes out of the total number of businesses.
Correct VAT Refund	3.30%	The proportion of VAT refunds processed accurately and timely out of the total VAT refunds.
Correct Execution of Procurements	3.30%	The proportion of procurement processes executed correctly out of the total procurement processes.
Bank Payments Between Firms	1.70%	The proportion of bank transactions between firms out of the total expected bank transactions between firms.
Bank Payments to Employees	1.70%	The proportion of bank transactions to employees out of the total expected bank transactions to employees.
Presence of a Quality Control System	3.30%	The proportion of businesses or entities with a quality control system in place out of the total number of businesses or entities.
Bright Economy Proportion	3.30%	

Subtotal:	30%	
Employees	20.00%	
Primary Job By Contract	1.70%	The proportion of the labor force with a primary job secured by a formal employment contract out of the total labor force.
Primary Job - Full Remuneration Contract	1.70%	The proportion of the labor force with a primary job that receives full remuneration under a formal contract out of the total labor force.
Secondary Job By Contract	1.70%	The proportion of the labor force with a secondary job secured by a formal employment contract out of the total labor force.
Secondary Job - Full Remuneration Contract	1.70%	The proportion of the labor force with a secondary job that receives full remuneration under a formal contract out of the total labor force.
Printing Receipts for Goods	3.30%	The proportion of transactions for goods where receipts are printed out of the total transactions for goods.
Printing Receipts for Services	3.30%	The proportion of transactions for services where receipts are printed out of the total transactions for services.
Correct Declaration of Tax on Natural People's Income	6.70%	The proportion of individuals correctly declaring their income taxes out of the total number of individuals required to do so.
Subtotal:	20%	
TOTAL:	100%	

Table 1: Indicator Weights⁷ & Descriptions [10]

3.2 Assumptions

As the 2023 data will be available in late 2024 in the newest BECI report of BICA, this research paper, therefore, assumes that:

1. The indicators will remain the same and will not undergo any expansion or shrinkage (7 statistical and 11+7 sociological indicators).
2. The weighting distribution between indicators will remain the same, as evident in the 2018-2022 data sample (Shown in Table 1 and Charts 1-3 above).

Data analysis in the following section will utilize statistical theory, including calculating average growth rates, moving averages, and working with regression models and trendlines. To ensure a higher estimation accuracy for the 2023 BECI value, this paper will showcase different scenarios that will collectively contribute to creating an accurate data range. In the different scenarios presented in the following sections, we will expand our dataset back to 2015 and analyze the hypothetical situation where the global economic disturbances of 2020 due to the COVID outbreak are omitted. Having a range and using different fragments of the 2015-2022 dataset will allow us to construct multiple hypotheses about the estimated 2023 BECI value, considering the current and most recent global socioeconomic events.

⁷ Weights are sourced from the latest, 2022 BICA report on BECI. Explanations of indicators are based on the author's understanding.

3.3 Observations

In our study, it is important to acknowledge the reasons behind the annual change in BECI values. Several regulatory adjustments and steps were taken in 2015 and 2016 to improve control and streamline the operations of the National Revenue Agency (NRA) and the Customs Agency and to electronize their services. These actions increased the amount of public debt that was collected. Joint inspections conducted in 2018 by the NRA, the GIT, and the police served as the impetus for initiatives to curb unethical labor practices and expose financial infractions. However, labor relations were still improving among those with the lowest wages [7]. The Bulgarian economy faced two mutually opposing forces in 2020, 2021, and 2022 that had entirely different effects on the incentives and behaviors of economic agents. One aspect of the situation was the inclination to utilize fraudulent techniques to endure. 2020 saw a 4% decline in real GDP and a loss of jobs, with 2,212 thousand fewer persons employed in 2020 than there were in 2019. Many businesses faced severe financial difficulties due to the severe crisis, which caused certain economic activities to cease operations temporarily or permanently and resulted in staff layoffs. Labor market difficulties persisted in 2021 and 2022. Many detrimental factors have displaced the pandemic crisis's implications, which are especially pertinent in 2020. These factors include populism, the administrative minimum wage setting process, the ongoing uncertainty surrounding pension relationships, the ongoing disproportionate increase in the minimum wage in line with labor productivity and the average wage, etc. [8].

Since 2018, the “Informally Employed” indicator has no longer been considered when performing calculations belonging to the statistical component of BECI. Instead, in its place, BICA introduced two new indicators (“Hired By Contract” and “Declared Labor Resource”) with the aim of more precisely measuring the true economic impact of activities associated with labor and employment and avoiding a leakage in annual GDP measurements. The “Socially Secured Employed” indicator was also put into practice with a similar purpose.

4. Data Analysis & 2023 Prediction

Component Bar Chart: GDP Sourced From Bright vs Hidden Sector

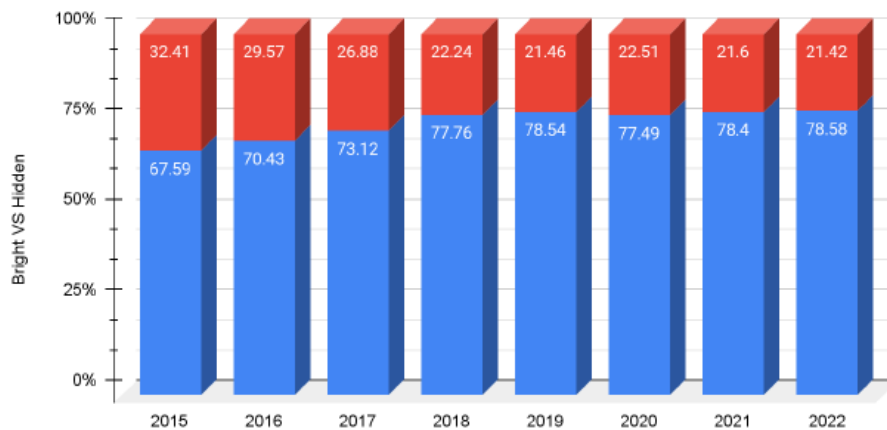


Fig. 3: GDP Sourced from Bright (blue) vs Hidden (red) sector

In the figure above, I have plotted the index over the last seven years, which shows the complementarity between the bright and hidden sectors. It can be seen that the bright economic contribution has been increasing over the years as the blue section of the graph has progressively risen. The red section of each bar, on the other hand, has been conversely experiencing reduction. Holistically, Fig. 3 showcases the process of “brightening” of the Bulgarian economy.

4.1 Introduction to Estimation Methods

In order to estimate the value of the Bright Economy Composite Index (BECI) in Bulgaria for 2023, this paper employs four distinct methods of estimation. Each method provides a unique approach to projecting the BECI value, offering a comprehensive range of possible outcomes. The goal is to create plausible upper and lower boundaries, representing ambitious and conservative calculations, respectively.

Each of these methods will be applied to the BECI data from 2015 to 2022. The analysis will include scenarios that account for global economic disturbances, such as the COVID-19 pandemic, by omitting data from 2020. This approach ensures a robust estimation process, allowing for multiple hypotheses about the 2023 BECI value.

By leveraging these diverse estimation methods, the paper aims to present a comprehensive and accurate forecast of Bulgaria’s 2023 BECI value. This multifaceted approach provides a detailed understanding of the potential future trajectory of Bulgaria’s bright economy, aiding policymakers and stakeholders in making informed decisions.

The methods outlined provide a solid framework for estimating the 2023 BECI value, considering various historical trends and economic factors. By applying these techniques, this paper aims to offer a reliable forecast that captures the complexities of Bulgaria’s bright economy.

4.2.1 Method 1: Average Annual Growth Rate (AAGR)

The first method involves calculating the average annual growth of the BECI from historical data in the 2015-2022 time frame. By analyzing the trend of indicator values⁸ over previous years, this method projects the 2023 value based on the average increase observed annually in the value. This approach assumes that past growth patterns will continue into the future, providing a straightforward yet effective means of estimation. The following formula is applied: *Average Annual Growth Rate* = $\frac{\text{latest value in dataset} - \text{first value in dataset}}{\text{number of years apart}}$. After an estimated value of annual growth is calculated for each indicator, it is added to the latest value in the dataset, resulting in an estimated 2023 indicator value.

$$2023 \text{ Estimated Value} = \frac{\text{latest value in dataset} - \text{first value in dataset}}{\text{number of years apart}} + \text{latest value in dataset.}$$

⁸ Working with the values instead of weighted values will not result in major changes to the final estimation because each indicator estimated value is then multiplied by the proportion that it contributes to the final weight.

Then, each projected indicator value is multiplied by its weight, resulting in a weighted value. All weighted values are then added and the total value represents the 2023 BECI estimate of 80.19⁹.

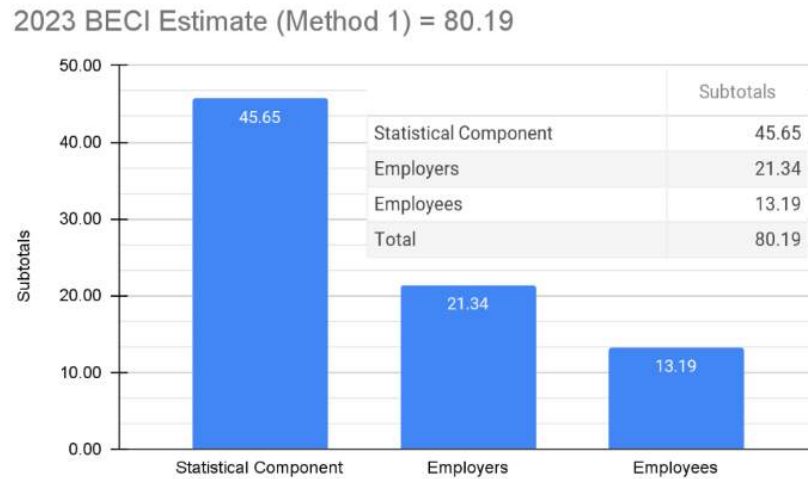


Fig. 4: 2023 BECI Estimate, Average Annual Growth, Whole Dataset

4.2.2 Method 2: Year-to-Year Average Growth Rate (YoY AGR)

The second method focuses on the year-to-year growth rates, and then results in an average annual growth rate from the beginning to the end of the dataset. Instead of a simple average over the entire period, this method examines the growth rate between each consecutive year: $Average\ Growth\ Rate = \frac{Ending\ value}{Beginning\ value} - 1$. By taking the average of these year-to-year changes, the method accounts for potential volatility and provides a nuanced projection that considers short-term fluctuations. After an AGR is calculated, the following formula is applied: $2023\ Estimated\ Indicator\ Value = (2022\ value) \times (1 + AGR)$. Each indicator value is then multiplied by its weight, which presumably remains constant, rendering a 2023 estimated weighted average. Similarly to other methods, the 2023 estimated weighted average for each indicator is combined, resulting in a cumulative 2023 BECI estimate. Using the whole dataset, the 2023 BECI estimate stands at 80.50, while in the scenario where only the most recent three years are taken into consideration, the 2023 BECI estimate stands at 79.41¹⁰.

⁹ For a detailed justification and explanation of calculations, please contact authors.

¹⁰ For a detailed justification and explanation of calculations, please contact authors.

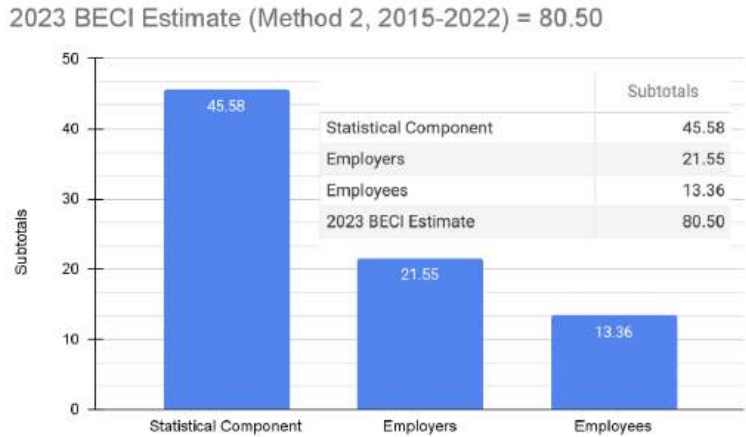


Fig. 5: 2023 BECI Estimate, Year-to-Year Average Growth Rate, Whole Dataset

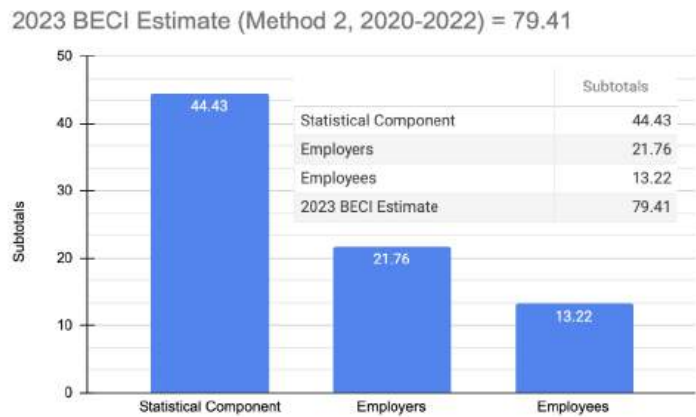


Fig. 6: 2023 BECI Estimate, Year-to-Year Average Growth Rate, 2020-2022 Data

Because using the whole dataset in this method renders a potentially overambitious BECI estimate of 80.50 as it incorporates a high growth rate from earlier years which inflates the average growth rate, our preferred estimate using this method is 79.41, which is the result of only considering the latest three years (2020-2022) in the dataset. In view of the mostly steady and consistent growth from the last five years (2018-2022), the second estimate showcased above appears to be more plausible. Nevertheless, the first estimate using this method, including data from the complete 2015-2022 time frame, will not remain disregarded in outlining the 2023 BECI range.

4.2.3 Method 3: Simple Moving Averages (SMA)

The third method employs moving averages to smooth out short-term variations and highlight longer-term trends. This technique calculates the average of the BECI values over a specified number of years, continuously updating as new data points are added. In this scenario, for calculating each average titled A_n , where “n” represents the number of averages considered in the final moving average calculation, an average of three years is computed. By using moving

averages, the method mitigates the impact of anomalous years and emphasizes the underlying trend in BECI values. Using this method absorbs extreme shocks that might have happened in the past, which in turn allows for variation and reduces extremities, while also involving simple regressions and nonlinear estimations. Because one indicator has no longer been in use after 2018, and two new ones have been introduced in its place, calculations using this method will only take into account values after 2018, where indicators remain unchanged. Averages are titled as follows:

$$A_1 = \frac{2018 \text{ value} + 2019 \text{ value} + 2020 \text{ value}}{3}$$

$$A_2 = \frac{2019 \text{ value} + 2020 \text{ value} + 2021 \text{ value}}{3}$$

$$A_3 = \frac{2020 \text{ value} + 2021 \text{ value} + 2022 \text{ value}}{3}$$

A_1 corresponds to an estimated 2021 BECI value, and similarly, A_2 coincides with the estimated 2022 BECI value, which means that A_3 stands for the predicted 2023 BECI value of 78.19.¹¹

Year	Actual Value	Moving Average	% difference
2020	77.50		
2021	78.39	77.79	0.77%
2022	78.56	78.17	0.50%
2023		78.19	

Fig 7: 2023 BECI Estimate, Simple Moving Averages, 2018-2022 Data

2023 BECI Estimate (Method 3, 2018-2022 Data)

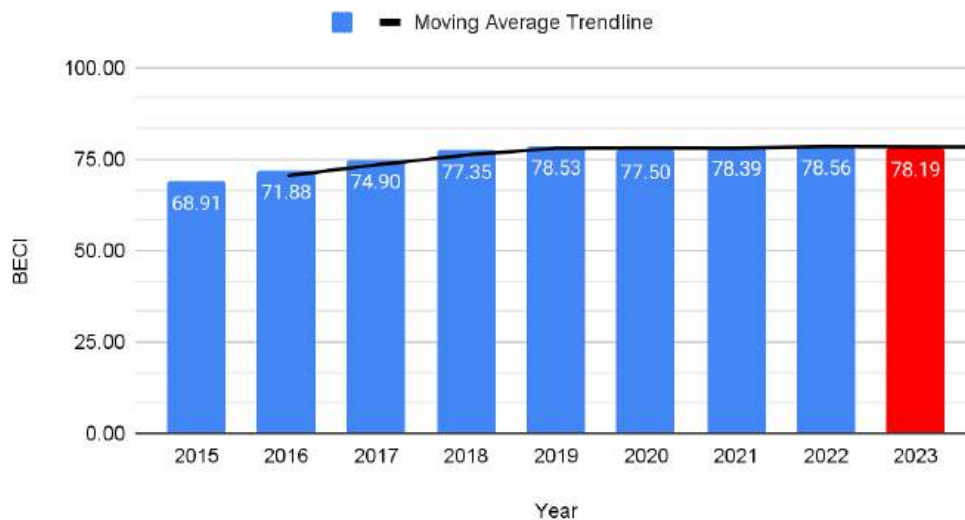


Fig. 8: 2023 BECI Estimate, Simple Moving Averages, 2018-2022 Data

¹¹ Detailed justification and explanation of calculations could be provided upon request.

4.2.4 Method 4: Regression Models and Trendlines (LR)

The fourth method utilizes regression models and trendlines to estimate the future BECI value. By fitting a statistical model to the historical BECI data, this approach identifies the relationship between time and BECI values. The model then projects the 2023 BECI value based on this identified trend. Regression analysis provides a rigorous framework for estimation, incorporating both linear and non-linear trends to capture the data's underlying dynamics.

Year	Time Period (x)	BECI (y)	Linear Regression Equation $y = mx + b + error$
2015	1	68.91	
2016	2	71.88	
2017	3	74.90	
2018	4	77.35	
2019	5	78.53	
2020	6	77.50	
2021	7	78.39	
2022	8	78.56	
2023	9	$\hat{y} = 81.60$	$y = 1.30x + 69.91 + error$ (based on the complete dataset)
		$\hat{y} = 78.75$	$y = 0.23x + 76.70 + error$ (based on last 5 years)

Table 4: Linear Regression Equations

2023 BECI Estimate (Linear Regression, Whole Dataset)

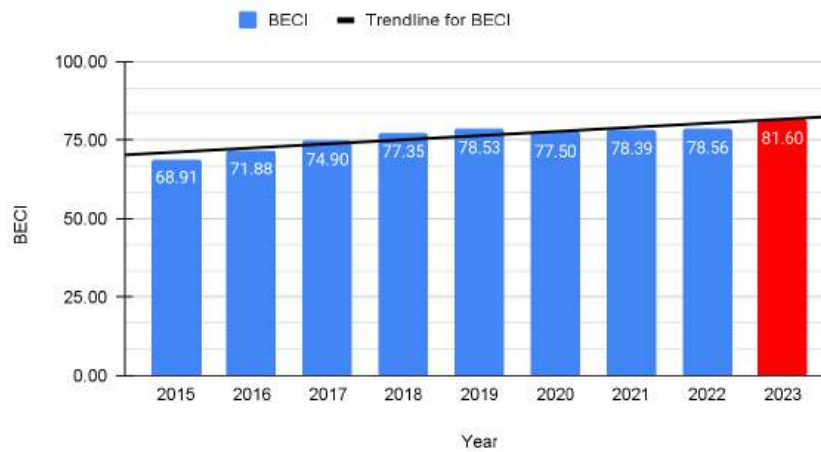


Fig. 9: Regression Trendline (2015-2023)

2023 BECI Estimate (Linear Regression, 2018-2022 Data)

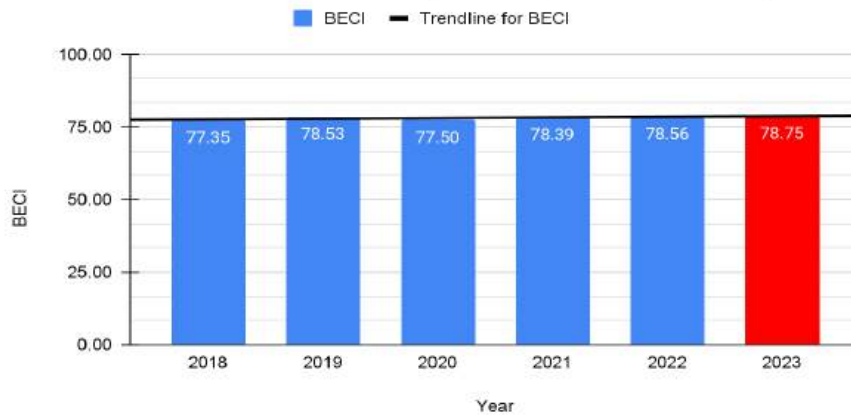
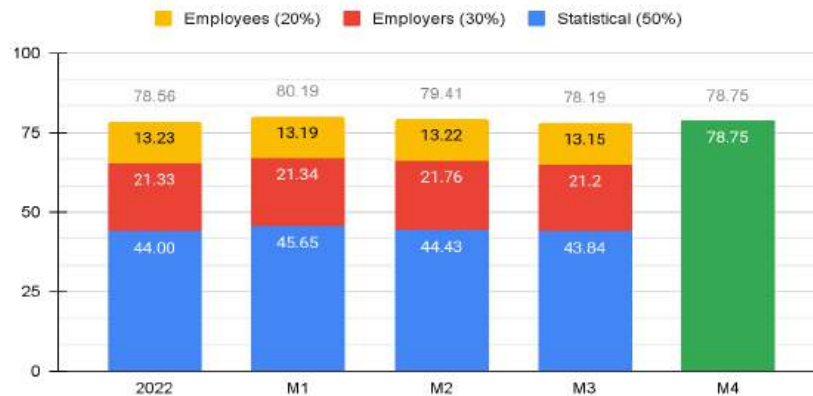


Fig. 10: Regression Trendline (2018-2023)

When taking into account data from the whole data set, the predicted value for the 2023 BECI is 81.60, but when taking into account data from just the last five most recent years, the predicted value for the 2023 BECI is 78.75. Our preferred estimate from this method is 78.75, using the equation $y = 0.23x + 76.70 + error$, because it more accurately observes the stickier and cohesive BECI growth as of 2018, as opposed to the fact that employing the whole dataset might inflate growth rates and lead to an overambitious 2023 BECI estimate. The predicted value of 81.60 seems unrealistic in view of BECI values over the last few years, consistently around 77-78, thus it will therefore not be considered in outlining the range in which BECI is predicted to fall.

5. Data Inference & Discussion



M1	M2	M3	M4
AAGR	YoY AGR	SMA	LR

Fig. 11: Intra Method Comparison

In the figure above, I decompose the BECI by subcategories for each of the methods employed. In this section, I will discuss different scenarios and look at how they compare across methods.

5.1 Sensitivity Analysis

Based on the comprehensive data analysis conducted for estimating the 2023 BECI value for Bulgaria, several inferences can be drawn regarding the effectiveness and implications of the different estimation methods. The estimated range for the 2023 BECI is between 78.19 and 80.50, reflecting the variability in outcomes depending on the methodology applied.

Method 3 (Simple Moving Averages), analyzing data from 2018 to 2022, yields the most conservative estimate at 78.19. By smoothing out short-term fluctuations, this method highlights the underlying stability but also tends to be more reserved in its projections, possibly due to the lack of major recent improvements in economic transparency. This estimate is significant because it suggests a cautious approach, acknowledging the slower, steadier progress in formalizing the economy. Due to changes in indicators in 2018, using data from only the last five years minimizes the risk of “losing” data values in converting averages between components.

The most optimistic estimate comes from Method 2 (Year-to-Year Average Growth Rate), which predicts the 2023 BECI at 80.50 after evaluating the complete 2015-2022 dataset. This figure represents the upper boundary of the estimated range and suggests a rather rapid improvement in economic transparency in the context of the 2022 BECI value. However, this method might potentially be slightly overly optimistic as it assumes that the strong growth seen in certain years will continue unabated without accounting for potential economic setbacks or stagnation that might appear as unforeseen. On the other hand, YoY AGR, which uses data from just the last three years, renders an estimate of 79.41. This figure is likely to be more closely aligned with what the actual BECI might be in 2023. The reason this method might be more accurate than working with AGR using values from the whole dataset is that it captures more recent trends while excluding older, more volatile data that may have been influenced by past economic policies, no longer causing any effects or changes in how indicators were measured. As a result, this method offers a balanced view, potentially reflecting the current state of the Bulgarian economy more accurately.

Method 1 (AAGR) produces a more ambitious estimate of 80.19. While this is still plausible, it is on the higher end of the spectrum. A potential issue with this method is that it includes years like 2015 and 2016, which had substantially lower bright economic proportions than the rest of the years in the dataset. This inclusion skews the annual growth rate upwards, giving the impression of higher growth than what might realistically be expected if considering only recent years. Since 2019, the growth has been more modest, suggesting that this method might potentially overestimate the 2023 BECI.

Focusing on Method 4, the linear regression models offer a nuanced perspective on estimating Bulgaria’s 2023 Bright Economy Composite Index (BECI). By analyzing the entire dataset from 2015 to 2022, the regression model projects a relatively high BECI value of 81.60,

which is also the most ambitious form of calculation in this research. This estimate reflects an assumption of long-term growth in economic transparency, suggesting that the bright economy in Bulgaria has improved abruptly compared to 2022, which raises questions about its plausibility. However, when the regression model is applied to a more recent timeframe, specifically the last five years, the projected value drops to 78.75. This lower estimate indicates that recent trends might not be as robust as the longer-term data suggests, possibly due to recent economic challenges or slower-than-expected improvements in transparency lately, potentially influenced by events such as the ongoing political conflicts worldwide or the COVID pandemic from a few years ago. This contrast highlights the regression method's sensitivity to the selected timeframe and underscores the importance of considering both long-term and short-term trends when making economic projections. While the overall regression approach is reliable for identifying general trends, it may overestimate the BECI when relying on older data or underestimate it if recent data reflects temporary slowdowns. Thus, while the regression models provide valuable insights, they should be viewed within the broader context of other estimation methods to avoid overly optimistic or conservative conclusions. In either case, it undoubtedly confirms the robustness of the different calculation methods and verifies the possible 2023 BECI range.

In all of the methods employed in the data analysis, based on the predictive power of the values for previous years, it could be concluded that all predictions are robust with a margin of error of about 5%, rendering a roughly 95% confidence interval.

5.2 Range Overview

The estimated BECI range of 78.19 to 80.50 implies that Bulgaria's hidden economy proportions would fall within the range of approximately 19.50 to 21.81. This indicates that while significant progress has been made, a considerable portion of economic activities remains outside the formal economy. When compared to European and global averages, Bulgaria still has a higher proportion of its economy in the shadows, signaling the need for continued efforts to reduce this hidden sector.

In conclusion, while the range of BECI estimates provides a comprehensive view of where Bulgaria stands, the YoY AGR method using 2020-2022 data appears to offer the most realistic projection for 2023 at an estimated 79.41. Continued efforts are necessary to decrease the hidden economy further, bringing Bulgaria closer to European and global standards.

5.3 Inferences About the Hidden Economy Sector

The estimated 2023 BECI range, projected to be between approximately 78% and 80.5%, provides critical insights into the state of Bulgaria's hidden economy. A higher BECI indicates a larger proportion of the economy operating within the formal, regulated sector, implying that the hidden economy -- comprising unreported income, informal employment, and other forms of economic activity that escape taxation and regulation-- is shrinking. Given this estimated range, the hidden economy will likely constitute roughly 19.5% to 22% of the total economy.

This level is significant as it suggests that while progress has been made in formalizing economic activities, a substantial portion of economic activity remains outside the official economy. Unlike European and global averages, where hidden economies typically range between 10% to 20% of GDP (with a European average of 17.3% [3]), Bulgaria’s estimated hidden economy is on the higher end, though not exceptionally so. The persistence of a relatively large hidden economy could indicate challenges in enforcing regulations, gaps in tax collection, or cultural factors that sustain informal economic practices.

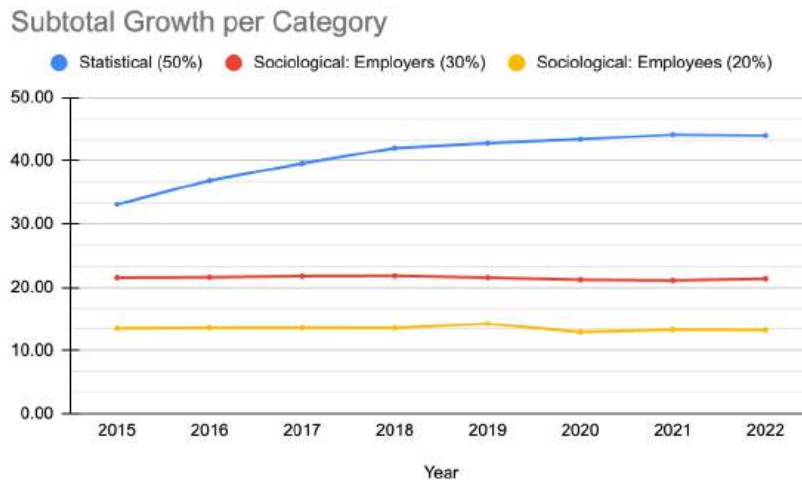


Fig. 12: Subtotal Growth, 2015-2022

I further deepdive into the analysis by looking into the subtotal trends by categories, as shown in Fig. 12. Based on the subtotal analysis using the trendlines, no notable variation within the subtotals could be observed. Hence, a slower and consistent BECI growth indicates that the statistical and sociological components are changing sluggishly. For example, because the regression method is only based on data from 2015-2022, while the estimate in aggregate shows that the bright economy proportion has gone up if we look at the subtotals more closely, we do observe that due to the COVID outbreak in 2020, the sociological component based on the employee responses actually reduced formal economic activities. We can observe a dip of approximately 1.5 units in the Employees subcategory, yet the change in the final value adds up to approximately 1 unit, which means that other indicators and subcategories can compensate for occasional or incidental dips in any of the categories, or in other words – that bigger changes in final BECI values might be set off across indicators. Nevertheless, 1 percent in the scale of GDP is indubitably a large number, which ends up being very costly to any government as it is not receiving tax for this said percent of the hidden economy. Therefore, we should really think of the subtotal values as crucial because they are directly tied to the scope of government expenditure.

It could be the case that the final BECI value is exactly the same over two different years, but the composition may have changed. This means that psychologically, people might have shifted towards adopting a brighter economy more than what is based on the actual activity

measured by the statistical component. In this case, it could be inferred that people generally start accepting a bright economy, and therefore, they are reporting higher employer satisfaction, for example, but when it comes to payment of VAT, for instance, we still see a reduction in contribution. Based on those observations, it may also be inferred that the sociological component greatly affects the acceptance of the indicators. In Bulgaria, however, there is a general lack of awareness among the public about the existence of BECI, which means many people miss out on the valuable insights it offers regarding the overall economy and specific sector activities. Among many potential reasons behind that, it is critical for us to acknowledge that when COVID occurred, the “Employees” sociological subcategory was reduced, possibly leading to people’s perception of the stability and security of formal employment being adversely affected. This could have contributed to a shift towards more informal or hidden economic activities as individuals and businesses sought to mitigate the uncertainties brought on by the pandemic.

That being said, to further reduce the size of the hidden economy, the government may need to implement policies focused on improving tax compliance, enhancing the effectiveness of regulatory bodies, and incentivizing businesses to operate transparently. Measures such as reducing the tax burden on small businesses, increasing digitalization in transactions, and enhancing public trust in government institutions could be effective in maintaining and potentially accelerating the decline in the hidden economy’s size. Additionally, incentivizing businesses to operate transparently, perhaps through tax breaks or subsidies for compliant enterprises, might encourage greater adherence to formal economic practices.

Holistically, although the BECI provides critical data on the distribution between the formal and informal sectors, helping to identify trends and inform policy decisions that could enhance economic transparency and stability, due to limited public knowledge, these insights remain underutilized, hindering their potential impact. Spreading more awareness about how BECI can be useful would empower businesses, policymakers, and citizens to make more informed decisions, ultimately contributing to a brighter and more sustainable economic future for Bulgaria.

6. Limitations

There were several notable limitations in the research:

6.1 Low Public Data Availability

A significant limitation in the research process was the low public accessibility of raw data used in the annual BECI calculations, sourced from governmental institutions. Information regarding the formulas used by BICA to develop the values for each indicator was unavailable to the mass public and non-disclosable to the organization, which did not allow us to execute calculations with true data, and we had to work with estimations accordingly. The preliminary sources for all data used in BECI calculations were also unavailable. In assembling the consolidated sheets with a complete list of values and weighted values in our data set, the values

in annual BECI reports by BICA did not always correspond to different BECI reports with data on the same period published on the same website. While BICA analysts go in retrospect and perform new calculations on BECI in previous years based on the emergence of new, more precise ways of measurement, yet older reports are still stored on the BICA webpage. Combined with the generally low data availability on this topic, the data collection component of this research, in turn, was significantly elongated.

6.2 Public Disengagement and Policy Efficacy

Over the recent decades, the Bulgarian government has tried to implement numerous policies to tackle the issue of a large proportion of GDP coming from hidden economic activity. Some of those include increasing the number of checks for indirect tax for businesses, which would lead to more effective monitoring; posing VAT restrictions on sellers; attempting to tackle this issue on an administrative level, which, in the best case scenario, would reduce corrupt practices. Throughout the years, one of the most crucial objectives for the government and other federal institutions has been to monitor whether businesses operate in accordance with the way they have documented their activity, as many employees are forced to work overtime, and their contracted labor does not coincide with the true amount of work they invest. The implementation of the majority of the policies has been relatively unsuccessful, and specialists claim that this is accounted for by the mass unawareness and disinterest in getting to know the depths and dimensions of normative practices within business sectors. With mass citizen disinterest in engaging, institutional efforts could never be truly effective in mitigating this issue and decreasing the share of GDP coming from the hidden sector. All of this being said, it is extremely difficult to pinpoint changes in the GDP distribution between the bright and the hidden sectors to any specific policy that the government may have implemented, which in turn makes it harder for professionals to back up claims on which types of policies are effective and should be continually put into practice.

6.3 Bias Introduced by Learning Over Time

Given that BICA uses a constant panel of respondents every year in an attempt to supply data belonging to the sociological component of the BECI, it is plausible to assume that respondents might exhibit learning over time. Potentially, this could lead to an overinflated bright economy proportion, which may not accurately represent and assess the true economic status quo in Bulgaria. Respondents may choose to underreport or overreport their activity belonging to any of the indicators that BECI is composed of, which means that a practical risk of data fabrication and manipulation exists. Since data availability was low and the sociological component of the BECI is primarily reliant on questionnaires and surveys, we were compelled to exclude consideration of the potential risk of data fabrication.

6.4 Lack of Transparency In Data Collection

In the scope of this research, we could not find the questions that the fixed panel of respondents in the sociological component is asked. Hence, it is important to acknowledge that there could be framing effects that might influence the value of the index year over year.

6.5 Scope of Chosen Methods

Potential limitations of the four methods of estimating the 2023 BECI value have been discussed in detail in section 5. The primary limitation that significantly impacted the research process remains low data availability, as discussed in section 6.1.

7. Conclusion

In this paper, we have conducted a thorough analysis to estimate Bulgaria's BECI value for 2023. By utilizing a combination of statistical estimation, scenario analysis, and trend extrapolation, we aimed to provide an accurate forecast of BECI, a key indicator of the size of the formal economy relative to the total economy. Our analysis was grounded in a detailed review of existing data, supplemented by projections and scenarios to address the inherent uncertainties in measuring the hidden economy.

The results indicate that the BECI for Bulgaria is likely to fall within the range of 78.19 to 80.50, suggesting that the hidden economy still constitutes approximately 20% of the country's GDP. This finding is consistent with previous years' trends, signaling both progress and the need for ongoing vigilance. The methodological approach adopted here was chosen over more complex econometric models due to its ability to handle data limitations and its applicability in scenarios where precise data is scarce. The combination of these methods provides a reliable and practical estimation, which is crucial for informing policy decisions. If more comprehensive and granular data were available, the analysis could have been enhanced by estimating the BECI using actual figures rather than relying on projections and scenario-based estimates. This would have allowed for a more precise and potentially more accurate understanding of the hidden economy's dynamics. However, the current approach offers a solid foundation for policymakers, even as it acknowledges the limitations posed by data availability.

That being said, there is still a great scope for future research related to the present findings. Further studies might consider BECI in other economic sectors beyond those mentioned in the current paper, which might include healthcare, education, or the agricultural sector. Additionally, cross-country comparisons of BECI could be valuable in shedding more light on how national policies against the hidden economy work in practice. Future research in this area may be oriented to constructing prediction models that include BECI, among other macroeconomic indicators, for establishing and assessing the trend of economic activity and hence improving policymakers' preparedness to respond to emerging challenges in this field or unforeseen economic circumstances. Given the fast development pace of digital technology, future research could also investigate how and to what extent digital transformation influences the size and nature of the shadow economy, especially concerning potential implications for the accuracy and relevance of the applied BECI measurements.

Works Cited

- [1] “Композитен индекс „Икономика на светло“”. АИКБ, 2021,
[https://www.ikonomikanasvetlo.bg/index.php?section=INT&aid=90&subid=105&lang=.](https://www.ikonomikanasvetlo.bg/index.php?section=INT&aid=90&subid=105&lang=)
- [2] “Hidden economy”. Oxford Reference,
[https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095935349.](https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095935349)
- [3] Schneider, F. “New COVID-related results for estimating the shadow economy in the global economy in 2021 and 2022”. Research Gate, May 2022,
[https://www.researchgate.net/figure/Size-of-the-shadow-economy-of-31-European-countries-in-2022-in-of-off-GDP-Source_fig3_361775720.](https://www.researchgate.net/figure/Size-of-the-shadow-economy-of-31-European-countries-in-2022-in-of-off-GDP-Source_fig3_361775720)
- [4] “Национален център Икономика на светло: Мисия”. АИКБ, 2021,
[https://www.ikonomikanasvetlo.bg/index.php?section=INT&aid=95&subid=114&lang=.](https://www.ikonomikanasvetlo.bg/index.php?section=INT&aid=95&subid=114&lang=)
- [5] “Динамика на Композитния индекс „Икономика на светло“”. АИКБ, 2023,
https://bica-bg.org/file/440/Dinamika%20na%20KI_2022.pdf
- [6] Kyle, S., & Warner, A. “The Shadow Economy in Bulgaria.”, US Agency for International Development, 2001, https://pdf.usaid.gov/pdf_docs/Pnacq971.pdf.
- [7] “Доклад за актуализация на Композитния индекс „Икономика на светло" към месец юни 2017 г.”. АИКБ, 2018,
https://marinesupport.biz/aikb/advices_files/106-24-1613657581.pdf.
- [8] “Доклад за актуализация на Композитния индекс „Икономика на светло" към месец юни 2018 г.”. АИКБ, 2019,
https://marinesupport.biz/aikb/advices_files/106-24-1613657596.pdf.
- [9] “Доклад за актуализация на Композитния индекс „Икономика на светло" към месец юни 2019 г.”. АИКБ, 2020,
https://marinesupport.biz/aikb/advices_files/106-24-1613657603.pdf.
- [10] “Динамика на Композитния индекс „Икономика на светло" и под-индекса „Заетост на светло" през 2022 г.”. АИКБ, 2023,
https://bica-bg.org/file/439/Doklad_2022_KI_Kompoziten_Index-SP.pdf.

A Systematic Review of Past, Current, and Future Approaches to Treat Breast Cancer

By Frances Pappas

Abstract

With regard to the past 50 years, much progress has been made in the realm of breast cancer therapeutics. Breast cancer research has led to exceptional progress with regard to the comprehension of breast malignancies. Researchers, based on these discoveries, are continuously aiming to create therapies resulting in less toxic effects and increased efficiency. Stratification systems have improved greatly from the Columbia Classification System of the 1950s to the establishment of four major subtypes of cancer in 2013 have aided in diagnosis processes. Surgery and radiation therapy, as localized therapies, have become less intrusive and increasingly precise. There has been great movement towards targeted therapies that have less effects on healthy cells. Current research studies and trials have demonstrated a greater comprehension of the inherent biological traits of breast cancer cells and using these discoveries to benefit treatments. While in the past, a patient diagnosed with breast cancer would be faced with surgery and cytotoxic systemic treatments, currently systemic treatments targeting a patient's specific subtype of cancer are universal treatment methods.

Introduction

Despite remarkable advancements in the treatment, understanding, and prevention of breast cancer over the past 50 years, it remains the second leading cause of cancer deaths among women, following lung cancer (“Breast Cancer Statistics”). Documentation of breast cancer dates back over 3,500 years to the Ancient Egyptians, who noted cases in medical papyri. The early recognition of breast cancer is said to have been caused by tumors being much more visible when compared to internal cancers (Mandal).

According to the American Cancer Society, 42,250 women are projected to die from breast cancer in 2024. However, the mortality rate has significantly declined since 1950, dropping from 31.9 to 19.3 per 100,000 (“Deaths by Breast Cancer”). This decline can be attributed to increased breast cancer awareness, routine screening, and high-quality treatments for both early and advanced stages (Jatoi et al. 251). Despite these improvements, breast cancer remains prevalent, with 1 in 8 women expected to develop the disease during their lifetime (“Breast Cancer Risk”). Disparities persist, as evidenced by the higher mortality rate among African American women (26.8 per 100,000) compared to Caucasian women (19.4 per 100,000) in the US (“Cancer of the Breast”). These disparities are due to delays in seeking treatment, lack of access to early detection and screening, and diagnoses at more advanced stages (“Breast Cancer Risk in American Women”)

Researchers and clinicians continue to explore improved methods of treatment, detection, and prevention, particularly focusing on disparities. For example, encouraging bi-annual mammograms for African American women could help address these disparities (“New Risk Model”). That being said, beyond expanding disease awareness and addressing socio-economic

disparities, the foundation of what saves lives are the therapeutics. This review will provide an overview of the history of breast cancer therapeutics, beginning with the field's initial attempts to treat the disease, all the way to cutting edge benchside research.

Past Therapeutics

In the mid-1950s, cancer cells were observed circulating in the blood during surgical removal of tumors. At that time, surgery was the sole therapy available. It was not until the discovery that chemotherapeutic agents, which are cytotoxic and preferentially target rapidly dividing cells, could kill circulating tumor cells in animal experiments. This led to the hypothesis that adjuvant chemotherapy would reduce tumor recurrence and improve survival rates in breast cancer patients (Fisher 1963.)

Chemotherapy was found to induce remission and, in some cases, caused advanced-stage breast cancers to disappear entirely (Donegan 6) With the advent of chemotherapy being included in the standard of care for breast cancer patients, there was a noticeable improvement in survival rates. Data from the End Results Section of the National Cancer Institute indicate that the five-year survival rate for breast cancer patients increased from 53 percent in 1940-1949 to 60 percent in 1950-1959. This improvement was largely due to advancements in treatment. Since 1950, there has been a significant decrease in the number of patients receiving surgery as their sole treatment and a substantial increase in those receiving primary therapy with hormones, radiation, and chemotherapy (Silverberg 2). Efforts to improve early detection of breast cancer have also been significant, as early-stage cancer is generally more treatable than later-stage disease.

In the 1950s, Dr. Haagensen, a prominent breast surgeon in the United States, strongly advocated for breast self-examination. At that time, women were often diagnosed with breast cancer only when tumors had grown large enough to be deemed untreatable, as mammography had not yet been developed (Thorton 2118) Additionally, Haagensen created the Columbia Clinical Classification (CCC) system in the early 1950s, which became the most prominent system for stratifying patients into stages. The CCC stages, denoted by letters A, B, C, and D, were used to determine appropriate surgical interventions, diagnoses, and predict post-treatment survival rates (Sakorafas, G.H., and Michael Safioleas 145).

Between 1915 and 1942, Haagensen treated 1,544 patients and correlated his findings with various clinical outcomes such as skin edema, ulcers, tumor fixation, and arm edema. The CCC stages demonstrated a significant spread in 5 year survival rates: 73% for stage A, 58% for stage B, 24% for stage C, and 0% for stage D. Based on these observations, the CCC stages were defined as follows: Stage A shows no grave signs of the presence of cancer cells within lymph nodes. Stage B shows no grave signs but possesses presence of cancer cells within mobile axillary lymph nodes less than 2.5 cm in size. Stage C demonstrates one of five grave signs. Stage D. Every other aforementioned stage with metastases (Nachlas 681).

In addition to the CCC stages, physicians looked for physical indications of the disease. Common symptoms included arm lymphedema, radiation-induced sarcomas, and chest

disfigurement, which were often indicative of breast cancer. In addition to the CCC stages, physicians looked for physical indications of the disease. Common symptoms included arm lymphedema, radiation-induced sarcomas, and chest disfigurement, which were often indicative of breast cancer (Lukong 64).

Building on improved classification and qualitative visual examinations, there was a push to include more quantitative and scientific technologies. Raul Leborgne, a leading figure in breast cancer screening, developed the technique of breast compression to reduce radiation exposure and produce more accurate and visible mammographic images. Before the advent of breast cancer screening in 1982, the incidence of breast cancer had remained stable. However, when mammographic screening became common practice for breast cancer detection, the incidence of breast cancer rose by 4% each year from 1982 to 1986, continuing to rise until 1991. This became the first year when mortality rates for breast cancer declined. Post 1991, the incidence rate of breast cancer continued to rise to a rate of 141.4 cases per 100,000 women. Between 1999 and 2003, the incidence rate began to decline and then stabilized at 127 cases per 100,000 women (Verdial 517). Since 2008, breast cancer incidence rates have now risen by over 20 percent (“Breast Cancer Statistics And Resources”). The increasing prevalence of mammograms resulted in a higher cancer detection rate. This revealed the common occurrence of the disease, motivating more research.

In addition to advancements in screening and chemotherapy, by the mid-19th century, Robert McWhirter of Edinburgh proposed the concept of combining simple mastectomy with radiotherapy. By 1963, this treatment method began gaining traction in Europe due to its association with less disfigurement. McWhirter conducted a study showing that cases eligible for radical mastectomy, when treated with simple mastectomy and radiation, resulted in a five-year cure rate of 62%. In contrast, those treated with radical mastectomy had survival rates between 35% and 45%. McWhirter also observed that arm edema, a common complication following radical mastectomy, rarely occurred after simple mastectomy and radiation therapy (Sakorafas, G.H., and Michael Safioleas 145). These early advancements in breast cancer therapeutics, from the observation of circulating tumor cells and the introduction of chemotherapy to the development of less invasive surgical techniques, laid the foundation for modern approaches to treatment.

Modern Therapeutics

In 1896, George Beatson Thomas was the first to delineate a connection between breast cancer and estrogen produced in the ovaries (Flach and Wilbert R43.) Recognizing that breast cancer is dependent on estrogen highlighted the need to decrease estrogen production in patients. This led to the development of the drug tamoxifen, which reduced the reliance on surgical treatments. This drug, as an antiestrogen, works to inhibit the binding of estrogen to estrogen receptors in the nucleus by competitively binding to the receptors. Thus, the growth of cancer cells is slowed. In 1977, the United States Food and Drug Administration approved the use of tamoxifen for patients with advanced breast cancer (Osborne).

Due to high rates of breast cancer cases and deaths of women in specific families, researchers were motivated to find the root of this susceptibility. In 1990, Mary-Claire King published a report hypothesizing the existence of a gene passed through genetics that could predispose individuals to breast and ovarian cancers. This report led to significant reforms in breast cancer procedures and diagnoses. Gilbert M Lenoir, a year later, validated King's findings. King named this theoretical gene BReast CAncer 1 (BRCA1). It was found that BRCA1 was implicated not only in hereditary breast cancer cases but also in early-onset breast cancer. This overgeneralization suggested the existence of another genetic mutation contributing to hereditary and early-onset breast cancers. Shortly thereafter, the BRCA2 gene was identified on chromosome 13q (Burachik et al.)

Following the identification of BRCA1, there was a push to isolate and clone the gene, which was accomplished by 1994. Genetic studies revealed that BRCA1 functions as a tumor suppressor gene. When BRCA1 loses its function, it leads to the growth of cancer cells (King 1462). BRCA1 and BRCA2 work together to protect cellular genes from damage during DNA replication by repairing double-strand DNA breaks (Roy et al 68). Identifying the BRCA1 or BRCA2 gene mutation at an early stage, can lead to timely treatment and diagnosis and a survival rate comparable to breast cancer patients not possessing a genetic mutation (King 1462).

In 1984, researchers discovered the neu gene in rat tumors. An ortholog in humans was later identified and given the name human epidermal growth factor receptor 2 (HER2.) The HER2 receptor plays a crucial role in the growth, repair, and division of epithelial cells, including those implicated in breast cancer. However, HER2 amplification leads to excessive receptor expression, promoting the growth of breast cancer associated with rapid cell proliferation and metastasis. Due to the aggressive nature of HER2-overexpressing breast cancers, a more aggressive and tailored therapeutic approach is required, as tamoxifen alone is insufficient (Rubin and Yarden S3). Slowly, breast cancer drugs have improved. By 1998, herceptin, a targeted therapy for HER2+ patients, was approved by the FDA. Herceptin was found to extend the survival rate of breast cancer patients with metastatic tumors that amplify the HER2 receptor. Patients treated with Herceptin and chemotherapy exhibited slower cancer progression compared to those receiving chemotherapy alone (Stebbing et al. 287).

In 2013, four major subtypes of breast cancer were classified using immunohistochemistry. These subtypes were determined by the expression of the hormone receptors: estrogen receptor positive (ER+), progesterone receptor positive (PR+), human epidermal growth factor receptor positive (HER2+), and triple-negative (TNBC), which is distinguished by an absence of the aforementioned hormone receptors. Based on these factors, luminal A, luminal B, HER2-positive, and triple-negative, are the four major subtypes of breast cancer. Luminal A tumors are ER and/or PR positive and HER2 negative, typically associated with the best prognosis among the subtypes. Luminal B tumors are ER positive and may lack PR, with a worse prognosis than luminal A. HER2-positive tumors overexpress the HER2 receptor and lack ER and PR receptors, making them more aggressive and fast-growing. Triple-negative breast cancer (TNBC) lacks ER, PR, and HER2 receptors, is highly aggressive, often diagnosed

at later stages, and has a high relapse rate. Identifying the specific subtype of breast cancer is crucial for determining the most effective treatment plan for each patient (Orrantia-Borunda).

In 2017, Ogivri (trastuzumab-dkst) was the first biosimilar drug to be passed by the FDA. Ogivri is used to treat metastatic HER2 positive breast cancer as a biosimilar to Herceptin. This is revolutionary as it is unlike generic drugs as it is a copy of a biologic drug (Rathore 1). Drugs of a brand-name can be either synthetic, formed by a chemical process, or biological, formed from living organisms. Synthetic brand-name drugs can be reproduced into affordable generic drugs. However, biologics involve large and complex molecules, so they cannot be reproduced. As these biosimilars are formed from living sources, their ingredients are not identical to their branded equivalent. Biosimilars are adjacent enough to name-brand drugs that they can accomplish identical therapeutic and clinical outcomes. These biosimilar drugs take the burden off of cancer patients as they cost less (“Biosimilars and Generic Drugs”). This is monumental as 47% of cancer patients and survivors have had medical debt related to their cancer (“Survivor Views”).

In 2019, Enhertu became the first targeted treatment for metastatic, inoperable HER2-positive breast cancer approved by the FDA. In the DESTINY-BreastO4 phase 3 trial, 555 adult patients with metastatic or inoperable HER2-low breast cancer were enrolled. The study revealed that patients experienced a significantly longer progression-free survival of 9.9 months and an overall survival of 23.4 months. This trial demonstrated a promising future for Enhertu in treating metastatic HER2-low and HER2-positive breast cancer patients (Siddiqui). Enhertu is a HER2-directed antibody-drug conjugate. HER2 is an oncogene that, when activated, provides the cell with anti-apoptosis signals, leading to cancer cell development (Gutierrez and Stiff 55). Enhertu binds to the HER2 receptor in the cell and damages its DNA, leading to cell death and preventing cancer cell replication (Varghese, Ryan et al).

In 2020, Trodelvy (sacituzumab govitecan-hziy), a targeted treatment for patients with triple-negative metastatic breast cancer (mTNBC) who had previously received at least two other treatments for metastatic cancer, was approved by the FDA (Cardillo 43). Trodelvy is an antibody-drug conjugate that targets the protein Trop-2, which is overexpressed on the surface of most breast cancer cells and plays a role in tumor growth. When Trop-2 is targeted by this conjugate, the cell is killed (“New Data for Trodelvy”).

The phase three ASCENT trial evaluated the effects of Trodelvy compared to standard-of-care chemotherapy in patients with mTNBC. The results demonstrated that Trodelvy was superior to single-agent chemotherapy. For patients without brain metastases treated with Trodelvy, the median progression-free survival was 5.6 months, compared to 1.7 months for those treated with chemotherapy (Fala).

In 2023, Capivasertib combined with fulvestrant was approved for the treatment of breast cancer patients with hormone receptor-positive, HER2-negative metastatic or advanced cancer who had not responded to previous treatments (Nierengarten 835). In a phase II trial, 140 patients with metastatic ER-positive breast cancer were randomized to receive either fulvestrant and capivasertib or fulvestrant and a placebo. Patients treated with capivasertib had a median

progression-free survival of 10.3 months compared to 4.8 months for those receiving the placebo. The median overall survival was 26 months for the capivasertib group versus 20 months for the placebo group, with survival differences becoming apparent after 12 months (Jones et al 1005). Capivasertib is an AKT inhibitor. Although breast cancer is not typically characterized by AKT amplification, AKT functions as an oncogene that promotes growth in both healthy and cancer cells.

These advancements in targeted therapies and personalized medicine have significantly improved treatment outcomes for breast cancer patients. That being said, the field still has a long way to go and many lives to save, so researchers and clinicians are actively pushing the boundaries to discover new and more effective treatments.

Future Therapeutics

Clinically, there are many groundbreaking drugs in the pipeline. A combination of three drugs for metastatic HER2-negative breast cancer showed a promising response in a phase 1b trial at the Johns Hopkins Kimmel Cancer Center. Patients were treated with entinostat, nivolumab, and ipilimumab, aiming to improve the response to immune checkpoint inhibitors. These drugs block checkpoint proteins, which normally prevent excessive immune responses but can also inhibit the immune system's ability to kill cancer cells (“Immune checkpoint inhibitor”). The trial generated a 25% response rate in patients with advanced HER2-negative breast cancer and a 40% response rate in those with triple-negative breast cancer (Roussos Torres et al 1). These response rates are comparable to many drugs approved by the FDA in the past decade. A 2019 study assessed 85 cancer drugs approved based on their response rates and found a median response rate of 41%. Notably, 16% of these drugs had a response rate below 20%, 33% below 30%, and 47% below 40% (Chen et al 915).

Inavolisib is an oral, small molecule PI3K (phosphatidylinositol 3-kinase) inhibitor. Alterations in the PI3K pathway are observed in all breast cancer subtypes, where it plays a crucial role in tumor cell growth and survival and serves as an essential regulator in many cellular processes. A PI3K inhibitor works by suppressing this pathway, thereby inhibiting cancer growth (Yu et al 635). In the phase 3 INAVO120 trial, 325 patients with HR-positive, HER2-negative breast cancer were treated with either a placebo or inavolisib plus Ibrance and fulvestrant. The trial aimed to determine whether the inavolisib-containing regimen led to longer progression-free survival (“Genentech”). The results showed that the progression-free survival rate for the inavolisib-treated group was 15 months, compared to 7.3 months for the placebo group (Pozzi et al 1415).

In addition to the groundbreaking research already advancing through the clinical pipeline, a substantial body of pre-clinical research at the bench is driving our understanding of breast cancer forward. A subset of cancer cells, termed cancer stem cells, possesses the ability to self-renew and drive tumor growth, as evidenced by research and clinical trials (Pozzi et al 1415). Breast cancer stem cells have been observed to become enriched and differentiate when treated with standard anti-breast cancer therapies. The mechanism behind this enrichment

remains unclear. Researchers from the Pajonk Lab at UCLA sorted non-breast cancer stem cells from patient samples and discovered that differentiated breast cancer cells, when exposed to radiation, transformed into induced breast cancer stem cells. These induced stem cells exhibited gene expression profiles similar to those of non-irradiated stem cells. The study concluded that radiation contributes to the increase in breast cancer stem cells following standard cancer treatments (Lagadec et al 833).

Breast cancer stem cells, like all cancer stem cells, exhibit high cellular plasticity. This plasticity allows these stem cells to readily adapt to targeted therapies, such as PI3K inhibitors, as well as cytotoxic therapies like radiation or chemotherapy. The side effect of radiation therapy that increases cancer stem cell populations is prompting clinicians to reassess its use for their patients, as radiation can lead to more aggressive and resistant cancers later in treatment. This side effect might be mitigated by adjusting the radiation dose and timeline or by incorporating additional drugs alongside chemotherapy. For instance, therapies that prevent the differentiation of cancer cells into cancer stem cells could potentially overcome this challenge.

At the Drexel University Reginato Lab, Kruppel-like factor 8 (KLF8) was found to regulate breast cancer stem cells in triple-negative breast cancer (“KLF8 KLF Transcription Factor 8”). KLF8 is involved in DNA repair and is shown to contribute to therapy resistance when overexpressed. The mechanism of action of chemotherapy and radiation therapy are centered around damaging the DNA of a cell. This is done to such an extent that the cell dies. Thus, if the genes that repair the DNA are upregulated, it is met with resistance. KLF8 overexpression has been identified as a factor in the resistance of standard triple-negative breast cancer treatments, including chemotherapy and paclitaxel. Cancer stem cells are known to promote and sustain long-term tumor growth, and the upregulation of KLF8 further enhances tumor resistance to therapy.

In an attempt to downregulate KLF8, the researchers utilized the SUM159 breast cancer cell line and knocked down KLF8 expression using a stably expressed shRNA. Both shKLF8 and shScramble expressing SUM159 cell lines were injected into the mammary fat pads of mice. After six weeks, the control injection (shScramble) resulted in significantly larger tumors compared to the shKLF8 injection, underscoring the importance of KLF8 in breast cancer progression and the regulation of breast cancer stem cells. Conversely, overexpressing KLF8 increased the breast cancer stem cell population and enhanced the cells' resistance to paclitaxel and chemotherapy (Le Minh et al).

In a 2024 study, metformin, an anti-diabetic drug, was combined with targeted therapies such as trastuzumab for HER2-positive breast cancer, tamoxifen for ER-positive breast cancer, and an antibody against the ROR1 receptor for triple-negative breast cancer (TNBC). A subpopulation of the breast cancer cells was first identified, and the concentration of each drug required for each cell line was determined using a cytotoxicity assay. Following a period of incubation, the cells were exposed to these drug concentrations. The anticancer effects of the drugs, both individually and in combination, were then measured, focusing on cell migration, tumor growth, and metastasis potential. The results demonstrated that tamoxifen, when used in

combination therapy, led to a significant reduction in tumor size and a lower potential for metastasis. These findings suggest that metformin, when used in combination with targeted therapies, could enhance the effectiveness of breast cancer treatment (Mahmoudi et al 10). These examples of clinical trials and laboratory research illustrate just a fraction of the ongoing efforts to improve breast cancer treatment. As research continues to advance, the future of breast cancer therapeutics holds great promise, offering hope for more effective and personalized treatment options for patients worldwide.

Contributing Factors to Breast Cancer

Having explored the therapeutic landscape of breast cancer treatment, it is equally important to understand the risk factors and lifestyle contributions that lead to the development of breast cancer. By examining these factors, we can gain insights into prevention and early detection strategies that complement advancements in treatment. Per the SEER Cancer Statistics Review, a woman's risk of developing breast cancer increases significantly with age. In a woman's 30s, her chance of getting breast cancer is approximately 1 in 204. This risk increases to 1 in 65 by her 40s. By the time she reaches her 60s, the risk rises to 1 in 28 (Howlander et al).

According to SEER Cancer Incidence Rates, African American and Caucasian women have the highest risks of breast cancer incidence in the US. Caucasian women have a 14% lifetime risk of developing breast cancer, partly due to factors such as having children at a later age, bearing fewer children, and using menopausal hormone therapy more frequently than other races. These factors are associated with an increased risk of breast cancer (Willett et al). African American women have a 12% lifetime risk of breast cancer (“Breast Cancer – 5-year age-adjusted”). Statistically, African American women are more likely to have children at a younger age and are less likely to breastfeed after childbirth compared to Caucasian women. Both of these factors are also linked to an increased risk of developing breast cancer (Pinheiro SP et al).

Inherited mutations in BRCA1 and BRCA2, genes involved in DNA repair pathways, are linked to a higher breast cancer occurrence. When these genes are mutated, cells have an increased risk of acquiring DNA damage, which can alter the expression of multiple genes and contribute to the onset of cancer. People of Ashkenazi Jewish descent have a greater risk of possessing BRCA1 and BRCA2 mutations than other populations, with about 1 in 40 Ashkenazi Jewish men and women carrying a BRCA1/2 gene mutation. In the general US population, these genetic mutations are rare, affecting approximately 1 in 400 to 1 in 800 individuals (“BRCA1 and BRCA2: cancer risks”).

Through the environment and food, many women are exposed to xenobiotics, which can disrupt the body's endocrine system. These endocrine disruptors create estrogen-like effects that influence a woman's risk of breast cancer. Many plastic additives found in water bottles, food containers, toys, plastic dishware, and more are classified as endocrine disruptor chemicals, such as BPA, BPF, and BPS (Stillwater et al). To lower the risk of developing breast cancer, women

can be physically active, maintain a healthy diet, get genetic testing, breast screenings and avoid plastic additives.

Conclusion

Breast cancer therapies have advanced remarkably in the past 70 years. Although chemotherapy, radiation therapy, surgery, and mammogram screenings are regarded as the standard of care for patients, many advancements in drugs and diagnoses have improved the mortality rate for US patients in the past. However, there are still a great number of lifestyle, environmental, and genetic factors which lead to high breast cancer incidence. Thus, the need for screening methods and therapies to combat incidence have arisen. As well, mortality rates for breast cancer have risen significantly over the past decade. Perhaps, by adjusting the aforementioned risk factors breast cancer incidence rates could decrease. The multitude of new therapy strategies to arise could make their way into clinical treatment methods. There must be more awareness with regard to the effects cancer therapies have on the body. By creating therapies with less toxic effects and risks of mortality could improve breast cancer patients' quality of life. Overall, this review highlights both the challenges and successes achieved in the field over the past decades. Significant progress has been made, and there is considerable optimism for future advancements in the therapeutic landscape.

Works Cited

- Breast Cancer Statistics | How Common Is Breast Cancer?
<https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html>. Accessed 12 June 2024.
- MD, Dr Ananya Mandal. "History of Breast Cancer." News-Medical, 11 Dec. 2009,
<https://www.news-medical.net/health/History-of-Breast-Cancer.aspx>.
- "Deaths by Breast Cancer U.S. Rate 1950-2019." Statista,
<https://www.statista.com/statistics/184615/deaths-by-breast-cancer-in-the-us-since-1950/>.
Accessed 13 June 2024.
- Jatoi, Ismail, and Anthony B. Miller. "Why Is Breast-Cancer Mortality Declining?" *The Lancet. Oncology*, vol. 4, no. 4, Apr. 2003, pp. 251–54,
[https://doi.org/10.1016/s1470-2045\(03\)01037-4](https://doi.org/10.1016/s1470-2045(03)01037-4).
- Breast Cancer Risk in American Women - NCI. 17 Dec. 2020,
<https://www.cancer.gov/types/breast/risk-fact-sheet>.
- "Cancer of the Breast (Female) - Cancer Stat Facts." SEER,
<https://seer.cancer.gov/statfacts/html/breast.html>. Accessed 13 June 2024.
- New Risk Model Aims to Reduce Breast Cancer Disparities in Black Women - NCI. 8 Dec. 2021, <https://dceg.cancer.gov/news-events/news/2021/breast-cancer-risk-calculator>.
- Fisher, B. "From Halsted to Prevention and beyond: Advances in the Management of Breast Cancer during the Twentieth Century." *European Journal of Cancer*, vol. 35, no. 14, Dec. 1999, pp. 1963–73, [https://doi.org/10.1016/S0959-8049\(99\)00217-8](https://doi.org/10.1016/S0959-8049(99)00217-8).
- Donegan, William. *History of Breast Cancer*. St. Louis. Saunders. 2002
- Silverberg, E., and A. I. Holleb. "Major Trends in Cancer:25 Year Survey." *CA: A Cancer Journal for Clinicians*, vol. 25, no. 1, Jan. 1975, pp. 2–7,
<https://doi.org/10.3322/canjclin.25.1.2>.
- Thornton, Hazel, and Raghu Ram Pillarisetti. "'Breast Awareness' and 'Breast Self-Examination' Are Not the Same. What Do These Terms Mean? Why Are They Confused? What Can We Do?" *European Journal of Cancer*, vol. 44, no. 15, Oct. 2008, pp. 2118–21, <https://doi.org/10.1016/j.ejca.2008.08.015>.
- Sakorafas, G. H., and Michael Safioleas. "Breast Cancer Surgery: An Historical Narrative. Part III. From the Sunset of the 19th to the Dawn of the 21st Century." *European Journal of Cancer Care*, vol. 19, no. 2, 2010, pp. 145–66,
<https://doi.org/10.1111/j.1365-2354.2008.01061.x>.
- Nachlas, Marvin M. "Irrationality in the Management of Breast Cancer. I. The Staging System." *Cancer*, vol. 68, no. 4, Aug. 1991, pp. 681–90,
[https://doi.org/10.1002/1097-0142\(19910815\)68:4<681::AID-CNCR2820680403>3.0.CO;2-2](https://doi.org/10.1002/1097-0142(19910815)68:4<681::AID-CNCR2820680403>3.0.CO;2-2).
- Lukong, Kiven Erique. "Understanding Breast Cancer – The Long and Winding Road." *BBA Clinical*, vol. 7, June 2017, pp. 64–77, <https://doi.org/10.1016/j.bbaci.2017.01.001>.
- Verdial, Francys C., et al. "Demographic Changes in Breast Cancer Incidence, Stage at Diagnosis and Age Associated with Population-based Mammographic Screening." *Journal of Surgical Oncology*, vol. 115, no. 5, 2017, pp. 517–22, <https://doi.org/10.1002/jso.24579>.

- “Breast Cancer Statistics And Resources.” Breast Cancer Research Foundation, <https://www.bcrf.org/breast-cancer-statistics-and-resources/>. Accessed 19 June 2024.
- Flach, Koen D., and Wilbert Zwart. “The First Decade of Estrogen Receptor Cistromics in Breast Cancer.” *Journal of Endocrinology*, vol. 229, no. 2, May 2016, pp. R43–56, <https://doi.org/10.1530/JOE-16-0003>.
- Osborne, C. Kent. “Tamoxifen in the Treatment of Breast Cancer.” *New England Journal of Medicine*, edited by Alastair J.J. Wood, vol. 339, no. 22, Nov. 1998, pp. 1609–18, <https://doi.org/10.1056/NEJM199811263392207>.
- Burachik, Natalia B., et al. “Discovery of BRCA Mutations: Historical Perspective of Its Scientific, Clinical and Social Impact.” *BRCA1 and BRCA2 Mutations - Diagnostic and Therapeutic Implications*, IntechOpen, 2023, <https://doi.org/10.5772/intechopen.108648>.
- King, Mary-Claire. “‘The Race’ to Clone BRCA1.” *Science*, vol. 343, no. 6178, Mar. 2014, pp. 1462–65, <https://doi.org/10.1126/science.1251900>.
- Roy, Rohini, et al. “BRCA1 and BRCA2: Different Roles in a Common Pathway of Genome Protection.” *Nature Reviews Cancer*, vol. 12, no. 1, 2012, pp. 68–78, <https://doi.org/10.1038/nrc3181>.
- Rubin, I., and Y. Yarden. “The Basic Biology of HER2.” *Annals of Oncology*, vol. 12, Jan. 2001, pp. S3–8, https://doi.org/10.1093/annonc/12.suppl_1.S3.
- Stebbing, J., et al. “Herceptin (Trastuzumab) in Advanced Breast Cancer.” *Cancer Treatment Reviews*, vol. 26, no. 4, 2000, pp. 287–90, <https://doi.org/10.1053/ctrv.2000.0182>.
- Orrantia-Borunda, Erasmo, et al. “Subtypes of Breast Cancer.” *Breast Cancer*, edited by Harvey N. Mayrovitz, Exon Publications, 2022, <http://www.ncbi.nlm.nih.gov/books/NBK583808/>.
- Rathore, Anurag S. “Approval of Ogivri.” *PDA Journal of Pharmaceutical Science and Technology*, vol. 72, no. 1, 2018, pp. 1–1, <https://doi.org/10.5731/pdajpst.2018.001105>.
- Siddiqui, Tasmiyah, et al. “Enhertu (Fam-Trastuzumab-Deruxtecan-Nxki) – Revolutionizing Treatment Paradigm for HER2-Low Breast Cancer.” *Annals of Medicine and Surgery*, vol. 82, Oct. 2022, <https://doi.org/10.1016/j.amsu.2022.104665>.
- Gutierrez, Carolina, and Rachel Schiff. “HER 2: Biology, Detection, and Clinical Implications.” *Archives of Pathology & Laboratory Medicine*, vol. 135, no. 1, Jan. 2011, pp. 55–62, <https://doi.org/10.1043/2010-0454-RAR.1>.
- Varghese, Ryan et al. “Fam-trastuzumab deruxtecan-nxki (Enhertu[®]): A narrative drug review.” *Cancer Research, Statistics, and Treatment*. Last modified December 2022. Last accessed August 31, 2024.
- Cardillo, Thomas M., et al. “Predictive Biomarkers for Sacituzumab Govitecan Efficacy in Trop-2-Expressing Triple-Negative Breast Cancer.” *Oncotarget*, vol. 11, no. 43, Oct. 2020, pp. 3849–62, <https://doi.org/10.18632/oncotarget.27766>.
- New Data for Trodelvy Demonstrate Clinical Efficacy Across Trop-2 Expression Levels in HR+/HER2- Metastatic Breast Cancer. <https://www.gilead.com/news-and-press/press-room/press-releases/2022/12/new-data-for-trodelvy-demonstrate-clinical-efficacy-across-trop2-expression-levels-in-hrher2-metastatic-breast-cancer>. Accessed 28 May 2024.

- Chen, Emerson Y., et al. "An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate." *JAMA Internal Medicine*, vol. 179, no. 7, July 2019, pp. 915–21, <https://doi.org/10.1001/jamainternmed.2019.0583>.
- Fala, Loretta. Trodelvy (Sacituzumab Govitecan-Hziy) a New Targeted Therapy for Patients with Metastatic Triple-Negative Breast Cancer. Aug. 2020, <https://www.oncpracticemanagement.com/special-issues/2020-oncology-drug-coding-and-updates/1695-trodelvy-sacituzumab-govitecan-hziy-a-new-targeted-therapy-for-patients-with-metastatic-triple-negative-breast-cancer>.
- Nierengarten, Mary Beth. "FDA Approves Capivasertib with Fulvestrant for Breast Cancer." *Cancer*, vol. 130, no. 6, Mar. 2024, pp. 835–36, <https://doi.org/10.1002/cncr.35238>.
- Jones, Robert Hugh, et al. "Capivasertib (AZD5363) plus Fulvestrant versus Placebo plus Fulvestrant after Relapse or Progression on an Aromatase Inhibitor in Metastatic ER-Positive Breast Cancer (FAKTION): A Randomized, Double-Blind, Placebo-Controlled, Phase II Trial." *Journal of Clinical Oncology*, vol. 37, no. 15_suppl, May 2019, pp. 1005–1005, https://doi.org/10.1200/JCO.2019.37.15_suppl.1005.
- "<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/immune-checkpoint-inhibitor>." NciAppModulePage, February 2, 2011.
- Roussos Torres, Evanthia T., et al. "Entinostat, Nivolumab and Ipilimumab for Women with Advanced HER2-Negative Breast Cancer: A Phase Ib Trial." *Nature Cancer*, Feb. 2024, pp. 1–14, <https://doi.org/10.1038/s43018-024-00729-w>.
- Yu, Miaomiao, et al. "Development and Safety of PI3K Inhibitors in Cancer." *Archives of Toxicology*, vol. 97, no. 3, 2023, pp. 635–50, <https://doi.org/10.1007/s00204-023-03440-4>.
- ClinicalTrials.Gov. <https://clinicaltrials.gov/study/NCT04191499>. Accessed 31 Aug. 2024.
- Genentech: Press Releases | Friday, Dec 8, 2023. <https://www.gene.com/media/press-releases/15012/2023-12-08/genentechs-inavolisib-combination-reduce>. Accessed 28 May 2024.
- Pozzi, Valentina, et al. "Cancer Stem Cell Enrichment Is Associated with Enhancement of Nicotinamide N-methyltransferase Expression." *IUBMB Life*, vol. 72, no. 7, 2020, pp. 1415–25, <https://doi.org/10.1002/iub.2265>.
- Lagadec, Chann, et al. "Radiation-Induced Reprogramming of Breast Cancer Cells." *Stem Cells (Dayton, Ohio)*, vol. 30, no. 5, May 2012, pp. 833–44, <https://doi.org/10.1002/stem.1058>.
- KLF8 KLF Transcription Factor 8 [Homo Sapiens (Human)] - Gene - NCBI. [https://www.ncbi.nlm.nih.gov/gene/11279#:~:text=a%20novel%20role%20and%20mechanism,invasion%20in%20gastric%20cancer%20cells.&text=Kruppel%2Dlike%20factor%208%20\(KLF8,essential%20for%20tumor%20cell%20proliferation](https://www.ncbi.nlm.nih.gov/gene/11279#:~:text=a%20novel%20role%20and%20mechanism,invasion%20in%20gastric%20cancer%20cells.&text=Kruppel%2Dlike%20factor%208%20(KLF8,essential%20for%20tumor%20cell%20proliferation). Accessed 6 June 2024.
- Le Minh, Giang, et al. "Kruppel-like Factor 8 Regulates Triple Negative Breast Cancer Stem Cell-like Activity." *Frontiers in Oncology*, vol. 13, 2023, p. 1141834, <https://doi.org/10.3389/fonc.2023.1141834>.

- Mahmoudi, Ghazal, et al. "Enhancement of Targeted Therapy in Combination with Metformin on Human Breast Cancer Cell Lines." *Cell Communication and Signaling*, vol. 22, no. 1, Jan. 2024, p. 10, <https://doi.org/10.1186/s12964-023-01446-0>.
- Chen, Emerson Y., et al. "An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate." *JAMA Internal Medicine*, vol. 179, no. 7, July 2019, pp. 915–21, <https://doi.org/10.1001/jamainternmed.2019.0583>.
- "SEER Cancer Statistics Review, 1975-2017." SEER, https://seer.cancer.gov/csr/1975_2017/index.html. Accessed 1 Sept. 2024.
- Themes, U. F. O. "Nongenetic Factors in the Causation of Breast Cancer." *Oncohemakey*, 9 July 2016, <https://oncohemakey.com/nongenetic-factors-in-the-causation-of-breast-cancer/>.
- SEER*Explorer Application. <https://seer.cancer.gov/statistics-network/explorer/application.html>. Accessed 1 Sept. 2024.
- Pinheiro, Simone P., et al. "Racial Differences in Premenopausal Endogenous Hormones." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, vol. 14, no. 9, Sept. 2005, pp. 2147–53, <https://doi.org/10.1158/1055-9965.EPI-04-0944>.
- BRCA1 and BRCA2 (PDQ®) - NCI. 3 Mar. 2023, <https://www.cancer.gov/publications/pdq/information-summaries/genetics/brca-genes-hp-pdq>.
- Stillwater, Barbara J., et al. "Bisphenols and Risk of Breast Cancer: A Narrative Review of the Impact of Diet and Bioactive Food Components." *Frontiers in Nutrition*, vol. 7, Nov. 2020, p. 581388, <https://doi.org/10.3389/fnut.2020.581388>.
- "Biosimilars and Generic Drugs in Cancer Care: What to Know." City of Hope, 8 July 2024, <https://www.cancercenter.com/community/blog/2024/07/biosimilars-generic-drugs-cancer>.
- "Survivor Views: Majority of Cancer Patients & Survivors Have or Expect to Have Medical Debt." American Cancer Society Cancer Action Network, 9 May 2024, <https://www.fightcancer.org/policy-resources/survivor-views-majority-cancer-patients-survivors-have-or-expect-have-medical-debt>.

The Origins of Multicellularity: Examining the Genetic and Environmental Factors Behind Multicellularity By Nakul Narayanan

Abstract

Multicellularity has allowed for the evolution of humans and animals, but researchers are still studying its origins. Given this, this research paper will discuss some genetic, environmental, and evolutionary drivers that may have contributed to multicellularity. It is posited that multicellularity evolved within the Choanoflagellate family, which split into modern Choanoflagellates and animals. Research on Choanoflagellates and other organisms has provided information on factors, including specific gene families and pathways, that could have contributed to the evolution of multicellularity. Through a thorough investigation of current research and literature, this paper examines genetic, environmental, and evolutionary contributors to multicellularity to better understand the field's current state. Overall, multicellularity likely resulted from a combination of genetic, environmental, and evolutionary factors. This includes the Wnt and Notch signaling pathways which allowed for cell differentiation and made energy usage in multicellular organisms more efficient. Understanding the evolution of multicellularity could allow for insight into the formation of Cancer and the Cambrian Explosion, which led to a massive diversification of life on Earth.

Introduction

The hand is a tool that humans have evolved that allows them to grip and hold things with precision. Without the hand, humans cannot use tools, and would not have progressed to the extent that they have currently. The hand is a product of multicellularity, which is the quality of being made of more than one cell. Multicellular organisms most likely split off from early single-celled organisms. Before multicellularity, organisms had to carry out all their functions within a single cell and were limited in size (Hinman and Cary). Multicellular organisms transcended those limits and moved onto land, eventually evolving into the multicellular animals that are seen today. Multicellularity did so by allowing organisms to designate specific roles to different cells, which has increased the efficiency of those cells and organisms. These roles allowed there to be different types of cells within the same organism, such as skin cells, and blood cells, which have differing functions. Skin cells are meant to stave off external threats to the organism such as diseases, ultraviolet light, chemicals, and injuries (Yousef et al.). Blood cells, on the other hand, are responsible for moving nutrients and oxygen around the body (Barbalato and Pillarisetty). However, both skin and blood cells are parts of the hand, which means the cells, although different, communicate and cooperate with each other to form a single object.

Historically, there have been many theories on the origin of multicellularity, with some scientists believing that it evolved from Ciliates, a unicellular eukaryotic group of organisms (Abraham et al.). Others believe that it evolved from an ancestor that displayed the behavior of Amoebas, a group of protists that can move with cell extensions called pseudopodia (Kang et

al.). The Amoeboid theory was formulated by Ernst Haeckel based on his now-obsolete theory of recapitulation, which theorized that the evolution of animals mirrored the development of those same organisms. He used this reasoning upon encountering a certain *Magosphaera planula*, which according to recapitulation fit perfectly the characteristics that an early ancestor of animals would need. He described *M. Planula* to be similar in division to the cells that made up sponge embryos, thus representing what should have been the earliest ancestor of animals. However, it should also be mentioned that *M. Planula* has not been seen since Haeckel's encounter with it. The Ciliate theory was developed by William Saville-Kent upon seeing similarities between small animals and ciliates. To defend this, he needed to prove that sponges alone were descendants of choanoflagellates, seeing that Haeckel had counted sponges as animals. Choanoflagellates are a group of single-celled organisms distinguished by the presence of a collar complex, which is a flagellum inside a feeding collar that helps in the capturing of bacteria. Choanoflagellates usually dwell in freshwater areas (Woznica). However, he also used an example species, *Proterospongia haeckelii*, as a perfect blend of sponge and choanoflagellate characteristics, to prove that there was an evolutionary connection between the two. *P. haeckelii* has also not been recorded since.

The leading theory today about the origin of multicellularity is that it evolved in the order Choanoflagellida, which then split off into the group including animals and current Choanoflagellates. The Choanoflagellate theory gained prominence due to improved molecular phylogenies showing that Choanoflagellates and Animals were closely related, supporting the theory that the two groups branched out from an early choanoflagellate ancestor. In addition, the discovery that Choanoflagellates could switch between Amoeboid and flagellate behavior supported earlier theories. Amoeboid behavior entailed the ability to move about the organism's corpus, and flagellate behavior necessitated the presence of a flagellum (Brunet and King). The leading theory on how multicellularity evolved is through a symbiotic relationship between cells, because of some combination of genetic, evolutionary, and environmental factors (Rose and Hammerschmidt). Therefore, this paper will investigate how these factors contributed to the evolution of multicellularity.

Section 1: Genetic contributors to multicellularity

Genetic Pathways are groups of genes that function together to execute steps of a biological process. Genetic changes are caused by errors in the maintenance and copying of DNA. For example, this may be caused by the DNA polymerase's failure to correctly read nucleotides and piece together copies of DNA. Some types of mutations include insertions and deletions, in which a nucleotide is added or removed, or missense and nonsense mutations, in which substituting nucleotides changes the amino acid the nucleotides code for (Suza and Lee). However, although changes in the genome are accidental, evolution selects for the development of genes that are best at preserving themselves, by benefiting the organism they belong to. Changes in genes and genetic pathways have influenced the evolution of multicellularity. For instance, the evolution of Cadherins and Integrins and the genes coding their construction might have led to the creation of the first multicellular organisms (Stevens et al.; Gul et al.). The

collaboration of the Wnt and Notch Signaling pathways could have allowed multicellular organisms to task cells with specific roles (Loh et al.). Meanwhile, mutations in the Homeobox genes of an organism could have coded for the development of more than one cell (Hubert and Wellik). These factors may have introduced the trait of multicellularity to previously unicellular organisms.

For an organism to be multicellular, its cells need to adhere to one other. Cadherins are a group of proteins present in cells that allow for cells to adhere to each other. In multicellular organisms, this allows for the retaining of the structure of tissues (Kaszak et al.). Integrins are a group of receptors that adhere to an extracellular matrix, in that way differing from Cadherins, which allow for Cell-cell adhesion. The extracellular matrix is present on the outside of the cell, and supports the interior portions, while aiding with the shaping of tissues and cell division (Takada et al.). The extracellular matrix is usually composed of water, proteins, and polysaccharides (Frantz et al.). Integrins are receptors able to sense forces acting upon the outside of the cell, being bound to the extracellular matrix (Takada et al.). Integrins also help form focal adhesions between cells, which are connections between the cells' cytoskeletons and the extracellular matrix surrounding them (Legerstee and Houtsmuller; Stevens et al.). Cadherins may have allowed early cells to adhere to each other during division instead of floating freely. However, for Cadherins to have contributed to the evolution of multicellularity, they should have been present at that time. The discovery of Cadherins in Choanoflagellates indicates that Cadherins might have been present in the first multi-celled animals as well, seeing that Choanoflagellates are single-celled, and are closely related to animals (Gul et al.). The combination of integrins and the extracellular matrix would have allowed early cells to adhere to each other if they were nearby, as the cytoskeletons of both cells would be connected to the same extracellular matrix.

The Notch Signaling pathway helps with the development of organs and tissues through a control of cell differentiation (Zhou et al.). It also allows for communication between cells, and this communication plays a part in deciding whether the cell is allowed to divide, or should die (Kopan and Ilagan). The Wnt pathway decides whether and when cells die, and controls cell division (Rim et al.). The pathway controls Cell Polarity, which is the distribution of organelles within the cell. Cell polarity is also a contributor to the shaping of the organism, and the development of their organs. However, the Wnt pathway also allows cells to divide, while giving their replicates a different phenotype (Loh et al.). Seeing that the Notch signal is short-ranged, two cells would have to be close to or in contact with each other in order for communication to occur (Kopan and Ilagan; Gazave et al.). The Wnt pathway may have contributed to the evolution of multicellularity because cells adhering to each other would be able to replicate with a different phenotype, allowing the Notch signaling pathway to designate those cells with different functions. This would help organisms meet both the characteristics needed from a multicellular organism, that being cells that adhere with each other and communicate despite having different phenotypes and roles (Loh et al.). Since the cells of a multicellular organism

must communicate with one other, the combination of Notch signals and the Wnt pathway might have contributed to the formation of the first multicellular organisms.

Homeobox (Hox) Genes control the development of organisms, and may have contributed to the development of multicellularity in early cells. Hox genes are usually present in animals due to the head-to-tail axis, which it also codes for. (Hubert and Wellik). Seeing that multicellularity most likely evolved from Choanoflagellates, resulting in a split that created the modern Choanoflagellate and animal groups, the evolution of Hox genes in animals might have split the two groups. Animals would have evolved multicellularity, while Choanoflagellates would have stayed single-celled. A mutation in the Hox genes of a cell may have coded for the inclusion of two or more cells within the body of the organism rather than a single cell. Since multicellularity offered multiple evolutionary benefits to organisms in a single-celled environment, evolution may have selected for the gene to be developed further.

Section 2: Evolutionary contributors to multicellularity

Evolutionary contributors to multicellularity are the factors that allowed the genes that coded for multicellularity to be preserved after they first evolved (Hershberg). Evolution selects for beneficial traits, allowing organisms with those traits to survive, and a chance for the genes of that organism to be passed on. However, genes can also be passed on via genetic drift, which is a random change to the frequency with which traits occur in a population (Bach). Once the trait of multicellularity had been first introduced into an organism by mutation, multicellularity had to pose some benefit in order for its corresponding gene to be selected for by the environment (Hershberg). It was down to the evolutionary benefits of multicellularity to allow that organism to survive until it was able to pass on its genes, thus preserving the trait of multicellularity. Some of the benefits of multicellularity included increased size and mobility, cell differentiation, and adaptability.

Multicellularity allowed organisms to grow in size, due to the fact that now they were not limited by how large their single cell could grow anymore (Boundless). Before multicellularity, cells could not grow too large, as the ratio between their surface area and their volume would be too small. For a large cell, too many resources would have to be transported through the cellular membrane. Larger cells provided organisms with immunity from predation, as filter feeding, which was most likely the means of predation during the evolution of multicellularity, limits itself to small animals. Organisms that were now larger had protection from predation from other large cells (Herron et al.). From an evolutionary standpoint, this means that there is now more of a chance of the organism replicating, which means that the gene coding for multicellularity is kept. On another note, being multicellular meant having differentiated cells, which could perform different jobs. As a single cell was not performing every function at once, multicellular organisms were able to use less energy per cell. Additionally, since multicellular organisms were better able to sense and move towards resources, they had some advantages over single-celled organisms (Colizzi et al.). However, if multicellular organisms were outright evolutionarily favored over unicellular organisms, evolution would have driven most unicellular organisms

toward multicellularity, which has not occurred. The ability of single cells to adhere without specification to any other cell would have led to inter-cell conflict and evolutionary competition between the cells of the same organisms. Additionally, organisms that evolved multicellularity through adherence would have had cells evolving independently of each other according to natural selection, making it hard for the organism to evolve complex multicellular traits controlled by the same gene. These disadvantages would have led most unicellular lineages to not evolve multicellularity (Márquez-Zacarías, Conlin, et al.). Thus, although multicellular organisms gained advantages dealing with resource acquisition, organisms evolving multicellularity through cellular aggregation created a degree of cell conflict.

In addition to increased size and mobility, multicellularity enabled cells to diversify, gaining differing phenotypes that correspond to differing functions (Colizzi et al.). In the process of cell differentiation, cells that were originally meant to perform every role could now replicate, but the daughter cells would have differing phenotypes, thanks to the Wnt and Notch Signaling pathways (Loh et al.). Multicellular organisms that developed from the divisions of single cells, thus displaying clonal multicellularity, would have reduced conflict between cells inside the organism (Márquez-Zacarías, Pineau, et al.). Under these circumstances, the cells would work together to access resources as a group, while protecting the organism from predators and hostile individuals on the outside. Additionally, cell differentiation helped organisms by making their functions more efficient (Rueffler et al.). Cells that specialized in a single task were more efficient than cells that had to distribute their resources between multiple tasks. The sum of these different cells working together was more than what single-celled organisms could complete individually. Additionally, cells with different roles sometimes had a synergy with each other, allowing for efficiency to be increased further. Functions whose efficiency depended on the position of the factor controlling them would favor differentiation. Multicellular organisms could locate their specialized cells in different places, allowing for the development of specialized organs with different functions that worked more efficiently due to their position.

Multicellularity offers organisms an increased ability to adapt based on their surroundings, and an ability to survive in diverse environments (Pepper et al.). Ongoing cell differentiation within the tissues of multicellular organisms was able to suppress inter-cell conflict and reduce competition within tissues for resources, meaning that multicellular organisms were able to spot and move toward resources more efficiently than unicellular organisms without having to compete with itself for those resources. An example of adaptation to surroundings is the evolution of legs on land animals, allowing for movement on land. In a diverse environment, the larger and more developed multicellular organisms would be at an advantage. This is because they may have access to more resources than unicellular organisms. If an environment runs out of resources, multicellular animals would be able to move further to find more than unicellular ones due to their mobility. Also, they would be able to acquire those resources, and in different environments, ongoing cell differentiation would allow for a more efficient method of adaptation to the environment. The mobility and size offered by multicellularity coupled with the ability of cells to differentiate and adapt to the environment

allowed multicellular organisms to access more resources with less competition. Organisms exhibiting these traits may have survived, and the trait of multicellularity may have been selected for, allowing for its further development.

Section 3: Environmental contributors to multicellularity

Environmental factors are the facets of the environment at the time of multicellularity's evolution that contributed to its development (Briggs). After single-cells first developed the trait of multicellularity, biological life underwent the process of the Cambrian Explosion, which was a rapid diversification in animal life. The Cambrian Explosion returned a diverse selection of fossils which have no earlier forms, which signals rapid growth in diversity and complexity in the genomes of organisms. Due to a variety of abiotic and biological processes and factors, the period of time preceding the Cambrian Explosion saw several environmental changes (Smith and Harper). The Cambrian Explosion caused a boom in diversity through changes to some environmental factors including the amount of nutrients that were available, the levels of oxygen on earth, and the composition of seawater (Zhang and Shu). Environmental factors such as these allowed there to be evolutionary benefits to the trait of multicellularity.

During the Cambrian Explosion, oxygen levels rose on earth, which aided with the switch from an anaerobic environment to an aerobic environment. At the same time, organisms on earth underwent an increase in complexity and diversity (Yang et al.). This evolution of complexity may have been facilitated by the increase in oxygen, as the usage of oxygen to generate energy through respiration generated more energy than pathways that did not use oxygen (Stamati et al.). The increase in oxygen led to the trait of aerobic respiration being evolutionarily favored, and the fact that organisms had switched to aerobic respiration enabled the evolution of multicellularity. This allowed organisms to generate the energy they needed to differentiate their cells and grow larger and more complex. The evolution of multicellularity would have also resulted in the spread of nutrients and oxygenation (Lenton et al.). Large eukaryotic particles would have sunk to the bottom of the ocean and would have successfully moved the center of oxygen usage from the surface of the ocean to the ocean floor. This would have promoted the development of multicellularity further, seeing that higher oxygen levels allow for larger size and more energy.

Along with oxygenation and energy, increasing population density, and environmental stability helped multicellularity develop (Tang et al.). During the Cambrian Explosion, sea waters became subject to constant salinities, the concentration of salt in water. Population density and salinity cause the switch between unicellular stages and multicellular filaments in some organisms. Thus, constant salinity as well as an increasing population density would lead organisms to exhibit the switch between unicellularity and multicellularity. Exhibiting the switch in states is an intermediate step in the process of achieving total multicellularity rather than exhibiting it as part of a stage of a life cycle. Additionally, increasing population density would have been favorable to the development of multicellularity as it would take away the pressure on individual cells to find and reach resources before other cells (Vroomans and Colizzi). The

evolution of cell adhesion, causing cells to stick to one other, allows large groups of cells to move towards a single resource, which in turn allows single cells to focus on replication rather than resource acquisition, as now it is a group of cells finding resources rather than a single cell. The constant salinities in oceans during the Cambrian Explosion allowed some organisms to make the switch between unicellular and multicellular states. At the same time, increasing oxygen levels in a newly aerobic environment allowed for the usage of cellular respiration, generating more energy than anaerobic pathways and allowing for increased cellular differentiation.

Section 4: Conclusion

Multicellularity was most likely a complex process that resulted from a combination of many different factors, including genetic, evolutionary, and environmental determinants. The leading theory on the origin of multicellularity, that it originated within the Choanoflagellate family, which then split into the group evolving into modern day animals and Choanoflagellates, may have been supported by these factors (Brunet and King). Choanoflagellate cells with adhesion molecules may have stuck to each other and created the first multicellular animals. Meanwhile, the evolutionary benefits of multicellularity during the environmental changes of the Cambrian Explosion might have allowed those organisms to survive and pass on their traits.

Research on the origin of multicellularity may be relevant to cancer, as cancer deals with the development of multicellular organisms from single cells. The evolution of multicellularity allowed for an increased risk of cancer due to the networks of communication between cells of the same organism that had been established. Some of the molecules and pathways that had enabled multicellularity to evolve may have created vulnerabilities in multicellular organisms' genes that when mutated could lead to cancer. Research into the vulnerabilities caused by multicellularity might provide information as to some factors that contribute to cancer (Trigos et al.). Additionally, the tendency of multicellularity to reduce intracellular competition within an organism may provide insight on cancer, as cancer is generally perceived as the result of individual evolution within cells. Research into the failure of multicellular organisms to stop intercellular competition and evolution within an organism may provide information on the causes of cancer (Nedelcu). Information on multicellularity's origins may be used in Cancer research due to the parallels between the causes of cancer and factors that lead to the evolution of multicellularity.

Some areas for further study are the effects of oxygen on the development of multicellularity within early unicellular organisms (Priest). This research may be headed in the direction of providing insight on life and conditions on Earth before the Cambrian Explosion, which inhibited the evolution of multicellularity. There most likely was not a single cause for the Cambrian Explosion, which contributed to the evolution of multicellularity. It was most likely a combination of factors that caused the Cambrian Explosion to trigger. Thus, this research may also head towards providing insight into the environmental factors that allowed for diversity and complexity during the Cambrian Explosion (Priest). The combination of genetic, evolutionary,

and environmental factors contributed to the evolution of multicellularity. This has allowed for the diversity seen among organisms, at the scale at which they exist today.

Works Cited

- Abraham, Jeeva Susan, et al. "Techniques and Tools for Species Identification in Ciliates: A Review." *International Journal of Systematic and Evolutionary Microbiology*, vol. 69, no. 4, Apr. 2019, pp. 877–94. PubMed, <https://doi.org/10.1099/ijsem.0.003176>.
- Bach, Jean-François. "Genetic Drift in Mammals." *Anais Da Academia Brasileira De Ciencias*, vol. 91, no. suppl 3, Aug. 2019, p. e20190339. PubMed, <https://doi.org/10.1590/0001-3765201920190339>.
- Barbalato, Laura, and Leela Sharath Pillarisetty. "Histology, Red Blood Cell." StatPearls, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK539702/>.
- Boundless. "4.4: Studying Cells - Cell Size." *Biology LibreTexts*, 1 Jan. 2024, [https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/General_Biology_\(Boundless\)/04%3A_Cell_Structure/4.04%3A_Studying_Cells_-_Cell_Size](https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/General_Biology_(Boundless)/04%3A_Cell_Structure/4.04%3A_Studying_Cells_-_Cell_Size).
- Briggs, Derek E. G. "The Cambrian Explosion." *Current Biology*, vol. 25, no. 19, Oct. 2015, pp. R864–68. www.cell.com, <https://doi.org/10.1016/j.cub.2015.04.047>.
- Brunet, Thibaut, and Nicole King. "The Single-Celled Ancestors of Animals." *The Evolution of Multicellularity*, by Matthew D. Herron et al., 1st ed., CRC Press, 2022, pp. 251–78. DOI.org (Crossref), <https://doi.org/10.1201/9780429351907-17>.
- Colizzi, Enrico Sandro, et al. "Evolution of Multicellularity by Collective Integration of Spatial Information." *eLife*, edited by Raymond E Goldstein and Aleksandra M Walczak, vol. 9, Oct. 2020, p. e56349. *eLife*, <https://doi.org/10.7554/eLife.56349>.
- Frantz, Christian, et al. "The Extracellular Matrix at a Glance." *Journal of Cell Science*, vol. 123, no. 24, Dec. 2010, pp. 4195–200. PubMed Central, <https://doi.org/10.1242/jcs.023820>.
- Gazave, Eve, et al. "Origin and Evolution of the Notch Signalling Pathway: An Overview from Eukaryotic Genomes." *BMC Evolutionary Biology*, vol. 9, no. 1, Oct. 2009, p. 249. BioMed Central, <https://doi.org/10.1186/1471-2148-9-249>.
- Gul, Ismail Sahin, et al. "Evolution and Diversity of Cadherins and Catenins." *Experimental Cell Research*, vol. 358, no. 1, Sept. 2017, pp. 3–9. ScienceDirect, <https://doi.org/10.1016/j.yexcr.2017.03.001>.
- Herron, Matthew D., et al. "De Novo Origins of Multicellularity in Response to Predation." *Scientific Reports*, vol. 9, no. 1, Feb. 2019, p. 2328. www.nature.com, <https://doi.org/10.1038/s41598-019-39558-8>.
- Hershberg, Ruth. "Mutation—The Engine of Evolution: Studying Mutation and Its Role in the Evolution of Bacteria." *Cold Spring Harbor Perspectives in Biology*, vol. 7, no. 9, Sept. 2015, p. a018077. PubMed Central, <https://doi.org/10.1101/cshperspect.a018077>.
- Hinman, Veronica, and Gregory Cary. "The Evolution of Gene Regulation." *eLife*, vol. 6, May 2017, p. e27291. *eLife*, <https://doi.org/10.7554/eLife.27291>.
- Hubert, Katharine A., and Deneen M. Wellik. "Hox Genes in Development and Beyond." *Development (Cambridge, England)*, vol. 150, no. 1, Jan. 2023, p. dev192476. PubMed Central, <https://doi.org/10.1242/dev.192476>.
- Kang, Seunggho, et al. "Between a Pod and a Hard Test: The Deep Evolution of Amoebae."

- Molecular Biology and Evolution, vol. 34, no. 9, Sept. 2017, pp. 2258–70. PubMed Central, <https://doi.org/10.1093/molbev/msx162>.
- Kaszak, Ilona, et al. “Role of Cadherins in Cancer-A Review.” *International Journal of Molecular Sciences*, vol. 21, no. 20, Oct. 2020, p. 7624. PubMed, <https://doi.org/10.3390/ijms21207624>.
- Kopan, Raphael, and Ma. Xenia G. Ilagan. “The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism.” *Cell*, vol. 137, no. 2, Apr. 2009, pp. 216–33. PubMed Central, <https://doi.org/10.1016/j.cell.2009.03.045>.
- Legerstee, Karin, and Adriaan B. Houtsmuller. “A Layered View on Focal Adhesions.” *Biology*, vol. 10, no. 11, Nov. 2021, p. 1189. PubMed Central, <https://doi.org/10.3390/biology10111189>.
- Lenton, Timothy M., et al. “Co-Evolution of Eukaryotes and Ocean Oxygenation in the Neoproterozoic Era.” *Nature Geoscience*, vol. 7, no. 4, Apr. 2014, pp. 257–65. www.nature.com, <https://doi.org/10.1038/ngeo2108>.
- Loh, Kyle M., et al. “Generating Cellular Diversity and Spatial Form: Wnt Signaling and the Evolution of Multicellular Animals.” *Developmental Cell*, vol. 38, no. 6, Sept. 2016, pp. 643–55. ScienceDirect, <https://doi.org/10.1016/j.devcel.2016.08.011>.
- Márquez-Zacarías, Pedro, Rozenn M. Pineau, et al. “Evolution of Cellular Differentiation: From Hypotheses to Models.” *Trends in Ecology & Evolution*, vol. 36, no. 1, Jan. 2021, pp. 49–60. DOI.org (Crossref), <https://doi.org/10.1016/j.tree.2020.07.013>.
- Márquez-Zacarías, Pedro, Peter L. Conlin, et al. “Why Have Aggregative Multicellular Organisms Stayed Simple?” *Current Genetics*, vol. 67, no. 6, Dec. 2021, pp. 871–76. PubMed, <https://doi.org/10.1007/s00294-021-01193-0>.
- Nedelcu, Aurora M. “The Evolution of Multicellularity and Cancer: Views and Paradigms.” *Biochemical Society Transactions*, vol. 48, no. 4, Aug. 2020, pp. 1505–18. PubMed, <https://doi.org/10.1042/BST20190992>.
- Pepper, John W., et al. “Animal Cell Differentiation Patterns Suppress Somatic Evolution.” *PLoS Computational Biology*, vol. 3, no. 12, Dec. 2007, p. e250. PubMed Central, <https://doi.org/10.1371/journal.pcbi.0030250>.
- Priest, Greg. “Framing Causal Questions about the Past: The Cambrian Explosion as Case Study.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 63, June 2017, pp. 55–63. ScienceDirect, <https://doi.org/10.1016/j.shpsc.2017.04.003>.
- Rim, Ellen Youngsoo, et al. “The Wnt Pathway: From Signaling Mechanisms to Synthetic Modulators.” *Annual Review of Biochemistry*, vol. 91, June 2022, pp. 571–98. PubMed, <https://doi.org/10.1146/annurev-biochem-040320-103615>.
- Rose, Caroline J., and Katrin Hammerschmidt. “What Do We Mean by Multicellularity? The Evolutionary Transitions Framework Provides Answers.” *Frontiers in Ecology and Evolution*, vol. 9, Nov. 2021. Frontiers, <https://doi.org/10.3389/fevo.2021.730714>.
- Rueffler, Claus, et al. “Evolution of Functional Specialization and Division of Labor.”

- Proceedings of the National Academy of Sciences, vol. 109, no. 6, Feb. 2012, pp. E326–35. pnas.org (Atypon), <https://doi.org/10.1073/pnas.1110521109>.
- Smith, M. Paul, and David A. T. Harper. “Causes of the Cambrian Explosion.” *Science*, vol. 341, no. 6152, Sept. 2013, pp. 1355–56. science.org (Atypon), <https://doi.org/10.1126/science.1239450>.
- Stamati, Katerina, et al. “Evolution of Oxygen Utilization in Multicellular Organisms and Implications for Cell Signalling in Tissue Engineering.” *Journal of Tissue Engineering*, vol. 2, no. 1, Nov. 2011, p. 2041731411432365. PubMed Central, <https://doi.org/10.1177/2041731411432365>.
- Stevens, Adam J., et al. “Programming Multicellular Assembly with Synthetic Cell Adhesion Molecules.” *Nature*, vol. 614, no. 7946, Feb. 2023, pp. 144–52. PubMed, <https://doi.org/10.1038/s41586-022-05622-z>.
- Suza, Walter, and Donald Lee. *Genetics, Agriculture, and Biotechnology*. Iowa State University Digital Press, 2021. iastate.pressbooks.pub, <https://doi.org/10.31274/isudp.2021.113>.
- Takada, Yoshikazu, et al. “The Integrins.” *Genome Biology*, vol. 8, no. 5, 2007, p. 215. PubMed, <https://doi.org/10.1186/gb-2007-8-5-215>.
- Tang, Si, et al. “An Environmentally Induced Multicellular Life Cycle of a Unicellular Cyanobacterium.” *Current Biology*, vol. 33, no. 4, Feb. 2023, pp. 764-769.e5. ScienceDirect, <https://doi.org/10.1016/j.cub.2023.01.069>.
- Trigos, Anna S., et al. “How the Evolution of Multicellularity Set the Stage for Cancer.” *British Journal of Cancer*, vol. 118, no. 2, Jan. 2018, pp. 145–52. www.nature.com, <https://doi.org/10.1038/bjc.2017.398>.
- Vroomans, Renske M. A., and Enrico Sandro Colizzi. “Evolution of Selfish Multicellularity: Collective Organisation of Individual Spatio-Temporal Regulatory Strategies.” *BMC Ecology and Evolution*, vol. 23, no. 1, July 2023, p. 35. BioMed Central, <https://doi.org/10.1186/s12862-023-02133-x>.
- Woznica, Arielle. “What Choanoflagellates Can Teach Us about Symbiosis.” *PLOS Biology*, vol. 22, no. 4, Apr. 2024, p. e3002561. PubMed Central, <https://doi.org/10.1371/journal.pbio.3002561>.
- Yang, Dong, et al. “Reactive Oxygen Species May Play an Essential Role in Driving Biological Evolution: The Cambrian Explosion as an Example.” *Journal of Environmental Sciences (China)*, vol. 63, Jan. 2018, pp. 218–26. PubMed, <https://doi.org/10.1016/j.jes.2017.05.035>.
- Yousef, Hani, et al. “Anatomy, Skin (Integument), Epidermis.” StatPearls, StatPearls Publishing, 2024. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK470464/>.
- Zhang, XingLiang, and DeGan Shu. “Causes and Consequences of the Cambrian Explosion.” *Science China Earth Sciences*, vol. 57, no. 5, May 2014, pp. 930–42. Springer Link, <https://doi.org/10.1007/s11430-013-4751-x>.
- Zhou, Bingham, et al. “Notch Signaling Pathway: Architecture, Disease, and Therapeutics.” *Signal Transduction and Targeted Therapy*, vol. 7, Mar. 2022, p. 95. PubMed Central, <https://doi.org/10.1038/s41392-022-00934-y>.

Why Do Civilisations Collapse? Is Our Civilisation in Danger? By Yanjia Yang

A prevalent issue among descendants of Indigenous Canadian communities is that they do not know what their native names are (Vowel 2017). While there are more than 70 Indigenous languages in Canada today, most of them have fewer than 1,000 fluent speakers (Toth 2022). Their Indigenous identities were replaced by Anglophone culture, which disconnected the First Nations peoples from their ancestry (France 2021). The preservation of native culture is a common challenge for many Indigenous communities, including in the United States and Australia (Romaine 2002).

The experience of Indigenous Canadians is unique in that their civilization rapidly declined as a direct result of foreign conquest. However, civilizations that faded across history often reveal more complex and long-term influences. Toynbee (1974) proposed that the demise of civilizations results from their reaction to external influences. Diamond (2005), with a biological emphasis, also argued that the common cause of collapse across societies is their response to environmental problems. In either case, it is rarely possible to attribute the demise of flourishing communities to one cause; the interplay between external impact and internal unrest is often the ultimate challenge.

The starting point to discussing why civilizations collapse is having a concrete definition of “civilization.” Civilization is the manifestation of a culture. Hegelian historicism defines a multifaceted civilization using a unified dialectic, where the language system, government, labor divisions, religion, arts, and social organizations are condensed into a single organism (Gill 2007). This approach abstracts away the diverse experiences within a culture to focus on its broader stages of growth. However, as Frye (1974) pointed out, “The question is not whether a culture is an organism, but whether it behaves enough like one to be studied on an organic model” (6). This Spenglian approach to historical analysis addresses the potential complications of this analogy. By analyzing the collective periods of growth and decay from a macro perspective, we can zoom in on the causal relationships between external pressures and internal decomposition.

“Collapse” is a nuanced word that requires careful elaboration. Popular narratives of collapse often simplify the long-term change with an apocalyptic termination. For instance, Gill (2007) wrote about the Mayan civilization, “There was nothing they could do or could have done. In the end, the food and water ran out – and they died” (299). In reality, however, the Mayan civilization consisted of 60 to 70 states that each experienced separate periods of decay (Middleton 2017). Even as cities were abandoned during the Terminal Classic period (AD 750–1050), new cities like Chichen Itza, Mayapan, and Uxmal continued to be built (Milbrath 2003). The Mayan civilization continued to evolve even after the arrival of the Spanish in 1502 and still have 8 million descendants today (Middleton 2017). Contrary to popular belief, communities could adapt and evolve through collapse without disappearing. Rather than biological extinction, collapse should be understood through definitive archeological features,

including the fragmentation of political institutions, the disintegration of customs and language, and the breakdown of centralizing functions (Schwartz 2006).

The term “our civilization” poses an equivocal assumption that there is a singular civilization that could be generalized as “ours.” If a civilization is considered the manifestation of a culture, then each culture today warrants a civilization of its own. However, modern globalization has shifted the definition of civilization to a more all-encompassing one. As Spengler (2021) observed, “In all Civilizations the ‘modern’ cities assume a more and more uniform type” (108). Cross-cultural communication buttresses the stability and interdependence of different cultures today. While honoring the diverse customs, languages, and social structures around the world, this essay focuses on the shared broader trajectory of “our” modern civilization.

This essay argues that while the failure to respond to foreign conquests and ecological pressures have historically led to the collapse of civilizations, not one factor could have played a deterministic role independently. Modern technology has fundamentally changed the nature of civilizations and allowed cultures to become resilient together, but it has also led to critical challenges for some endangered cultures around the world.

Ecological Factors

Ellen Churchill Semple (1911) aptly wrote, “Man is a product of the earth’s surface” (1). In the case of the Mayans, their agricultural success was founded on the exploitation of woodlands. Without awareness of natural preservation, the Mayans used unsustainable cultivation methods that led to widespread deforestation and soil erosion (Diamond 2005). Over multiple centuries, their land gradually became vulnerable to deterioration and environmental disturbance.

While environmental exploitation led to issues for Mayans as it did for many other ancient civilizations, it was not as deterministically fatal as some historians contend (Middleton 2017). The impact of human activity on the environment at the time would be marginal compared to that of today. In the industrialized world, global warming is happening “at a rate not seen in the past 10,000 years” (NASA n.d.). Mass extraction, production, and pollution is changing the earth tens of thousands of times faster than the ancient cultivation of woodlands could. As unsustainable as Mayan agriculture may be, it still had limited impact on the wide expanse of the empire and was unlikely the primary cause of its collapse.

Another proposition focuses on natural disasters, suggesting that prolonged droughts across the Mayan states ultimately led to total collapse (deMenocal 2001). Stalagmite records show that “the most extreme droughts in the entire 3,300-year record” coincided with the classic Mayan crisis from AD 700 to 1135 (Webster 2007, 62). The three multi-century droughts in this period led to severe decreases in crop yields, which could have been aggravated by unsustainable farming practices and led to widespread famine.

However, it would be premature to identify a causal relationship between droughts and socio-political crises solely based on co-occurrence. As Maher et al. (2011) suggest, “current

evidence [of Mayan records] does not support simple cause-and-effect relationships between rapid climate-change events and cultural changes” (3). Humans were able to adapt to a wide range of environments even without modern technology. The late Bronze citadels of Mycenae and Tiryns continued for many decades even after earthquakes overtook the regions (Finné 2017). Another example would be how the droughts on the island of Anatolia led to improvements in water distribution systems, demonstrating the resilience of ancient civilizations (Jacobsen 2022).

We must apply caution when designating ecological and environmental factors as the cause of collapse. While nature is significant in shaping the foundation of ancient cultures, ecological problems on their own have unlikely wiped out an established civilization. Oftentimes, the social upheaval that precipitated from the ecological problems also played a major role in the disintegration of the political structure.

Foreign Conquest

Toynbee’s Challenge and Response theory (1974) posits that the fate of a civilization hinges on how it responds to external challenges. In the case of the Mayan civilization, ecological challenges like environmental decay and natural disasters were intertwined with internal unrest, which then reflected in the empire’s weak response against foreign threats. When faced with famine after the droughts, Mayan rulers sought to expand their territories to acquire new means of agricultural production. The conflicts between different Mayan states brewed into internecine warfare, which weakened the foundation of Mayan society (Cowgill 1979)..

When the Spanish conquistadors arrived three centuries later, the Mayan empire struggled to respond effectively due to pre-existing internal unrest (Tedlock 1993). Divided and competing over natural resources, the Mayans could not unite against the Spanish albeit outnumbering them. The threats of European military technology, infectious diseases, and warfare tactics overwhelmed the Mayan empire, leading to the gradual disintegration of its political structures, educational systems, and distinct lifestyle. The inability to defend against foreign conquerors underscores Toynbee’s proposition that a civilization’s demise often stems from its inadequate responses to external adversities.

The impact of foreign conquest was more direct in some cases, like the Fall of Rome. The replacement of the Western Roman rulers by barbarian kingdoms was one of the most well-documented political collapses of classical antiquity (Kyle 2016). Even then, the external threat was not an independent factor and was closely tied with persistent internal tensions. On one hand, major epidemics like the Plague of Cyprian (AD 250–270) annihilated the Roman population and made it difficult to defend the empire’s borders (Kyle 2016). Meanwhile, dissent between the Roman rulers exacerbated the plight, leading to weakened economy, corrupted military, and the separation of localized communities without loyalty to the empire (Watkins n.d).

Instead of being the sole destructors, the barbarians were only the last straw among the broader problems in Western Rome (Kyle 2016). Just as ecological challenges and internal

divisions made the Mayans vulnerable to foreign threats, the pre-existing biological crises and internal dissent in Rome must not be overlooked.

Current Civilization

In the Second Discourse, Rousseau (1754) posits that there is an irreconcilable conflict between man and nature. As early humans explored the wilderness, they inevitably disrupted the homeostasis of the earth in the process. The initial disturbance of the environment was necessary for civilizations to achieve discoveries and technological progress before they could exert reservation efforts (Bailey 2022). As in the case of the Mayans, their unsustainable farming methods were a necessary step of learning. Before having the means to understand the environmental consequences of their actions, a period of experimentation was required to establish the society's foundation.

However, the situation is vastly different today. As we shift the question to whether our civilization is in danger, the profound impact of human destruction cannot be ignored. Unlike the Mayans, we can understand our impact on the earth with science, yet we have not slowed down on environmental destruction. Our modern lifestyle is changing the earth faster than millennia before (NASA n.d.). While technology and globalization have enhanced humanity's ability to mitigate existential threats as a whole, certain parts of the global civilization are under disproportionate levels of ecological threats.

The Kiribati people face a particularly dire situation, as their 33 low-lying islands make them extremely vulnerable to climate change. The World Health Organization (2024) has encouraged the Kiribati government to enforce relocation policies due to rising sea levels and increasing climate instability. However, relocation comes at a significant cultural cost for Kiribatians, as they risk losing their language, religion, and customs. Under displacement, the Gilbertese language could face a similar fate as many of the endangered languages today. The inevitable assimilation that comes with relocation would effectively end the 4000-year-long civilization (Cauchi 2021). While technology makes our global civilization resilient to environmental challenges, the worsening climate crisis could erase ancient cultures and cause irreversible damage to humanity beyond technology's saving.

Human impact, whether environmental or political, is a significant cause of cultural erosion today. Indigenous peoples in Canada face similar challenges of assimilation as the Kiribatians. In the mid-nineteenth century, Prime Minister Sir John A. Macdonald led the forced removal of Indigenous peoples to reservations. Native Canadians were forced to assimilate into mainstream Canadian culture, as Indigenous children were placed into reformatory boarding schools similar to the Native Americans in the U.S. (Jacobs 2012). The lasting trauma and disruption among Indigenous communities is a result of global modernization.

Technology, globalization, and knowledge of past societies "put us at lower risk" than ancient civilizations (Diamond 2005, 17). However, the destructive power of technology and the heightened interdependence between different societies also put us at higher risk in some aspects. Although ancient civilizations may be more vulnerable to climate change, the ones that did

collapse likely suffered from other external threats and internal disturbances as well. Rather than dramatizing the failures of pre-modern societies, we must understand the complex causal relationships between different factors that led to the collapse.

Our civilization today is of a different nature than those before industrialization and globalization. As Diamond (2005) wrote, “We shouldn’t be so naive as to think that study of the past will yield simple solutions, directly transferable to our societies today” (17). We should leverage the technological advancements today to reconcile our relationship with the natural world as well as with each other.

Works Cited

- Bailey, Ryan. "Searching for an Environmental Self-Love: Humans and Nature in Rousseau's Discourse on the Origins of Inequality." University of Chicago (2022).
<https://doi.org/10.6082/uchicago.4336>.
- Cauchi, John P., Stefano Moncada, Hilary Bambrick, and Ignacio Correa-Velez. "Coping with environmental hazards and shocks in Kiribati: Experiences of climate change by atoll communities in the Equatorial Pacific." *Environmental Development* 37 (2021).
<https://doi.org/10.1016/j.envdev.2020.100549>.
- Cowgill, George L. "4 Teotihuacan, Internal Militaristic Competition, and the Fall of the Classic Maya" In *Maya Archaeology and Ethnohistory* edited by Norman Hammond and Gordon R. Willey, 51-62. New York, USA: University of Texas Press, 1979.
<https://doi.org/10.7560/750401-007>
- Diamond, Jared M. *Collapse: How societies choose to fail or succeed*. London: Penguin Books, 2011.
- Facing History & Ourselves, "Aggressive Assimilation," last updated October 29, 2019
<https://www.collectionscanada.gc.ca/obj/thesescanada/vol2/OKQ/TC-OKQ-7557.pdf>
- Finné, Martin, Karin Holmgren, Chuan-Chou Shen, Hsun-Ming Hu, Meighan Boyd, and Sharon Stocker. "Late Bronze Age Climate Change and the Destruction of the Mycenaean Palace of Nestor at Pylos." *PLOS ONE* 12, no. 12 (2017).
<https://doi.org/10.1371/journal.pone.0189447>.
- France, Paige. "A Lost Indigenous Identity - the Medium." *The Medium*, November 16, 2021.
<https://themedium.ca/a-lost-indigenous-identity/>.
- Frye, Northrop. "'The Decline of the West' by Oswald Spengler." *Daedalus* 103, no. 1 (1974): 1–13. <http://www.jstor.org/stable/20024181>.
- Gill, Richardson B., Paul A. Mayewski, Johan Nyberg, Gerald H. Haug, and Larry C. Peterson. "DROUGHT AND THE MAYA COLLAPSE." *Ancient Mesoamerica* 18, no. 2 (2007): 283–302. <https://doi.org/10.1017/S0956536107000193>.
- Government of Canada. "First Nations in Canada." May 2, 2017.
<https://www.rcaanc-cirnac.gc.ca/eng/1307460755710/1536862806124>.
- Hegel, Georg, Wilhelm Friedrich, and Duncan Forbes. *Lectures on the Philosophy of World History*. Translated by Hugh Barr Nisbet of Cambridge Studies in the History and Theory of Politics. Cambridge: Cambridge University Press, 1975.
- Jacobsen, Matthew J., Jordan Pickett, Alison L. Gascoigne, Dominik Fleitmann, and Hugh Elton. "Settlement, environment, and climate change in SW Anatolia: Dynamics of regional variation and the end of Antiquity." *PLOS ONE* 17, no. 6 (2022).
<https://doi.org/10.1371/journal.pone.0270295>.
- Middleton, Guy M. "Do civilisations collapse?" *Aeon*. Aeon Media Group, November 16, 2017.
<https://aeon.co/essays/what-the-idea-of-civilisational-collapse-says-about-history>.

- Milbrath, Susan and Carlos Peraza Lope. "Revisiting Mayapan: Mexico's Last Maya Capital." *Ancient Mesoamerica* 14, no. 2 (2003): 1-47.
https://www.academia.edu/3682158/Revisiting_Mayapan_Mexicos_Last_Maya_Capital._Susan_Milbrath_and_Carlos_Peraza_Lope_.
- NASA. "Evidence - NASA Science." <https://science.nasa.gov/climate-change/evidence/>.
- Rousseau, Jean-Jacques. *Discourse on the Origin of Inequality A Discourse on a Subject Proposed by the Academy of Dijon: What is the Origin of Inequality Among Men, and is it Authorised by Natural Law?* Translated by G. D. H. Cole. Zürich: ETH Zurich, 1754.
- Romaine, Suzanne. *Vanishing voices - the extinction of the worlds languages*. Oxford University Press, 2002.
- Semple, Ellen C. "Influences of Geographic Environment On the Basis of Ratzel's System of Anthro-Geography." *Nature* 88, no. 101 (1911). <https://doi.org/10.1038/088101a0>.
- Schwartz, Glenn M., and John J. Nichols, eds. *After Collapse: The Regeneration of Complex Societies*. University of Arizona Press, 2006. <http://www.jstor.org/stable/j.ctt181hxqb>.
- Spengler, Oswald and Charles Francis Atkinson. *The Decline of the West*. London: Arktos, 2021.
- Tedlock, Dennis. "Torture in the Archives: Mayans Meet Europeans." *American Anthropologist* 95, no. 1 (1993): 139–52. <http://www.jstor.org/stable/681183>.
- Toth, Katalina. "The Death and Revival of Indigenous Languages." *Harvard International Review*, January 19, 2022.
<https://hir.harvard.edu/the-death-and-revival-of-indigenous-languages/>.
- Toynbee, Arnold J. *A Study of History*. New York: Dell Publishing Co., Inc, 1974.
- Vowel, Chelsea. *Indigenous writes: A guide to first nations, métis & Inuit issues in Canada*. Newburyport: Portage & Main Press, 2017.
- Watkins, Thayer. "The Economic History of the Western Roman Empire." San José State University. Accessed July 10, 2024.
<https://www.sjsu.edu/faculty/watkins/barbarians.htm#:~:text=The%20Roman%20Empire%20during%20the,River%20at%20the%20Black%20Sea>.
- World Health Organization. "Countering the Climate Crisis in Kiribati." March 8, 2024.
<https://www.who.int/westernpacific/news-room/feature-stories/item/countering-the-climate-crisis-in-kiribati#:~:text=Kiribati%2C%20home%20to%20around%2012%2C000,the%20world%20to%20climate%20change>.

Teenage Addiction: How Social Media and the “Doom Scroll” has Contributed to Mental Ailment By Jesleen Walia

Abstract

Excessive consumption of any form can prove to be detrimental, and in this day and age, we face rising concerns over the excessive consumption of media and technology. This has previously been exemplified through research on Internet Addiction Disorder, which showed the negative effects of over-consumption of the internet and games on the human brain. Now, we face a new challenge, that of overconsumption and overexposure to various forms of media including social medias, online streaming platforms, and more. In consequence, as various social platforms broadly target the demographic of adolescent children and teenagers, we now see a rise in not only mental distress in the form of depression, anxiety, a reduced attention span, and more- but also in a neurological form. This includes a reduced ability to learn and process, as well as a larger predisposition to addiction or practicing addictive behaviors. This paper will overview the new algorithms of social media including the “doom scroll” phenomena and expand on how they introduce a new set of problems to the adolescent brain-especially that of addiction and an unnatural imbalance of certain neurotransmitters, including dopamine.

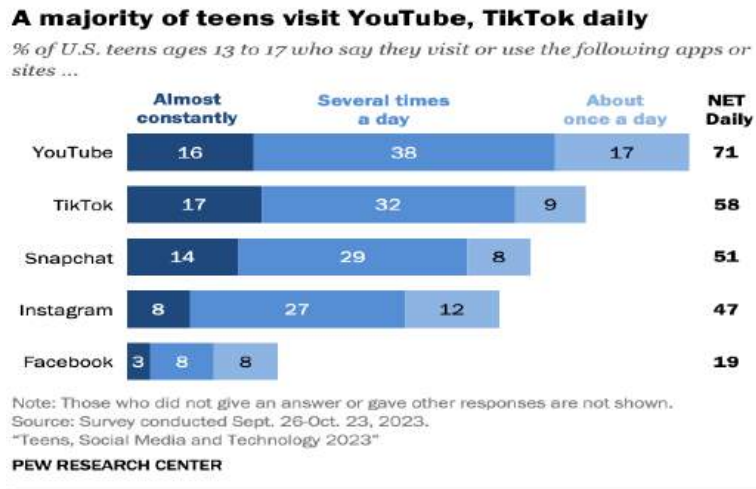
Introduction

In 1930s America, a period most historically categorized by the Great Depression, the economy fell into a recession. Thousands of dollars were lost, bank establishments fell, and a sense of mistrust and unease shrouded the working class. This not only left a concern for what the future of the country was to look like but also created a deep sense of emotional turmoil. Many families who relied on a singular income were left unable to support themselves, and as a result, in a desire to escape this harsh reality, emerged the sociological term, escapism.

Movies and theater were a newly surfaced concept of the time that only grew in popularity during the Great Depression as those who faced economic hardship began to “escape” their problems by means of cinema. Today, the term is still in use, perhaps even more so, with the emergence of social media and new technologies. While this new digital era has been defined by both its benefits and setbacks, one social issue that seems recurring and constant, is that of excessive teenage social media use as a form of escapism. This paper will analyze a current review of data on how teenage use of social media affects the adolescent brain, promotes addictive behaviors, and hinders brain development and neuroplasticity.

In 2023, Pew Research Center (Washington D.C, USA) conducted a study on teenage social media use, utilizing an online survey with 1,453 U.S. teenagers who were recruited through national random sampling. The survey was “weighted to be representative of U.S. teens ages 13 to 17 who live with their parents, by age, gender, race and ethnicity, household income, and other categories.” and discussed both social media use, and device ownership amongst teenagers (13-17). The study found that when asked if they use Tiktok or Instagram, 46 out of 99 teenagers reported using it “several times a day” and 14/99 reported using it “almost constantly.”

This “almost constantly” can also be defined by hours. As per Gallup’s 2023 Familial and Adolescent Health survey, over half (51%) of U.S. teens (13-19 years) spend a minimum of four hours daily on social media, at an average of 4.8 hours every day” and this steady increase of screen time usage can only raise the question, how does this neurologically affect the adolescent brain and its development?



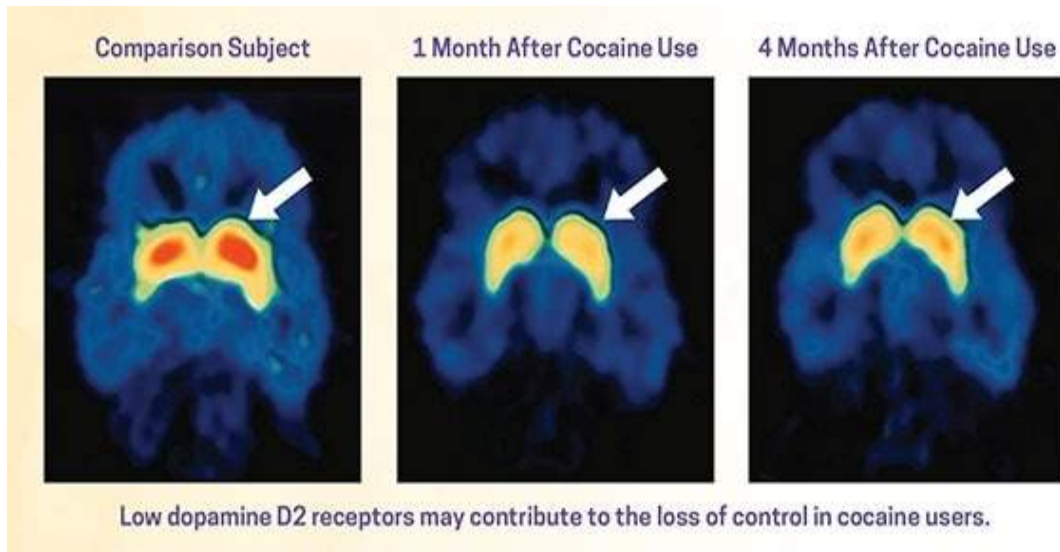
(Pew Research Center, 2023)

It has been previously found that social media has a negative correlation with mental health, especially within the margin of teenagers aged 13-17 who are “more predisposed to mental health issues” (American College of Obstetricians and Gynaecologists). In fact, upon such rising concerns about the effects of social media, the current United States Surgeon General, Vivek Murthy, issued an advisory warning claiming “the mental health crisis among young people is an emergency -- and social media has emerged as an important contributor,” calling for health warnings, statistics, and transparency from major social media companies on the effects of social media use (United States Department of Health and Human Services). Following this advisory there have also been numerous studies linking teen mental health with social media usage. A JAMA medical journal published their results of a study, finding- via a longitudinal household survey interview, that amongst “American teens ages 12-15, those who used social media over three hours each day faced twice the risk of having negative mental health outcomes, including depression and anxiety symptoms” (Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth, Jama Network, 2019). The Surgeon General's advisories are reserved for significant public health challenges that require the nation’s immediate awareness and action, and with such concern and call for action, the causation between excessive social media use and mental ailment becomes clear.

Dopamine is a neurotransmitter “most notably involved in helping us feel pleasure as part of the brain's reward system. Sex, shopping, smelling cookies baking in the oven — all these things can trigger dopamine release, or a ‘dopamine rush’” (Harvard Health Publishing). However, it is also involved in reinforcement. Dopamine is the defining part of your internal reward system- that is, when you get a surge of dopamine you often feel a sense of accomplishment, happiness, or influx of pleasure. However, there is a fine line, and when faced with high levels of dopamine throughout the day there is a higher risk of addiction and addictive behaviors similar to that of, or also demonstrated through, the feeling one may attain from consuming drugs. Additionally, it has been found that “a dopamine deficiency also contributes to a down mood... lack of motivation, and [loss of] concentration” symptoms commonly associated with depression. Now, how does social media come into play?

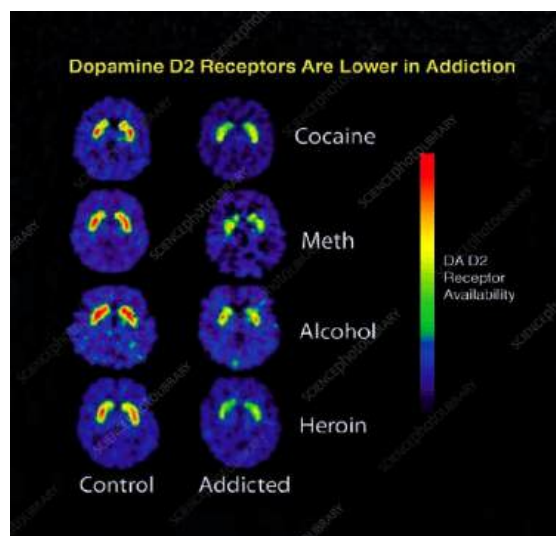
Take for example the popular social media platform Instagram. With each post comes the ability to gain likes, shares, reposts, and comments, and with each of these is a paired surge of dopamine. With an even stronger correlation, is Instagram reels, a new addition to the platform that features small 15 to 30-second videos in a vertical scroll layout. Similar to TikTok, this “added feature” has thousands of small videos in continuous recession without break or pause. The science behind this? Each reel gives the user a surge of dopamine, a surge so intense that it has been coined “artificial.” Now, if teenagers, whose brains are not yet fully developed, and who spend over 4 hours a day on such platforms, face such an intense surge of dopamine, their natural levels are thrown off balance, and dopamine receptors become fried. If each video provides a “hit” many will keep scrolling continuously to maintain that level of dopamine. In turn, with prolonged exposure to such an algorithm, one's “base levels” of dopamine are thrown off. Natural forms of dopamine such as trips, new experiences, social encounters, exercise, and more, no longer give the same feeling or amount of pleasure and satisfaction. This not only reduces one's ability to stay “grounded” in their personal experiences, but also leaves them going back for more- more reels, more TikTok's, and more social media to reach that level of dopamine again. Sounds familiar?

Addiction is defined as a “treatable, chronic medical disease involving complex interactions among brain circuits, genetics, the environment, and an individual's life experiences” and can include “behaviors that become compulsive and often continue despite harmful consequences” (American Society of Addiction Medicine, 2019). It is considered a psychological and neurological disorder as “it involves functional changes to brain circuits involved in reward, stress, and self-control.” (National Institute on Drug Abuse, 2020). With excessive use, social media can lead to dopamine addiction, defined perfectly by the new social media term, “doom scrolling.” This phenomenon where one scrolls through reels or TikTok videos continuously without realizing how much time has passed, exhibits a compulsive behavior that affects their brain structure in relation to their dopamine systems and feelings of reward. To provide a clearer image of just how dangerous social media is,

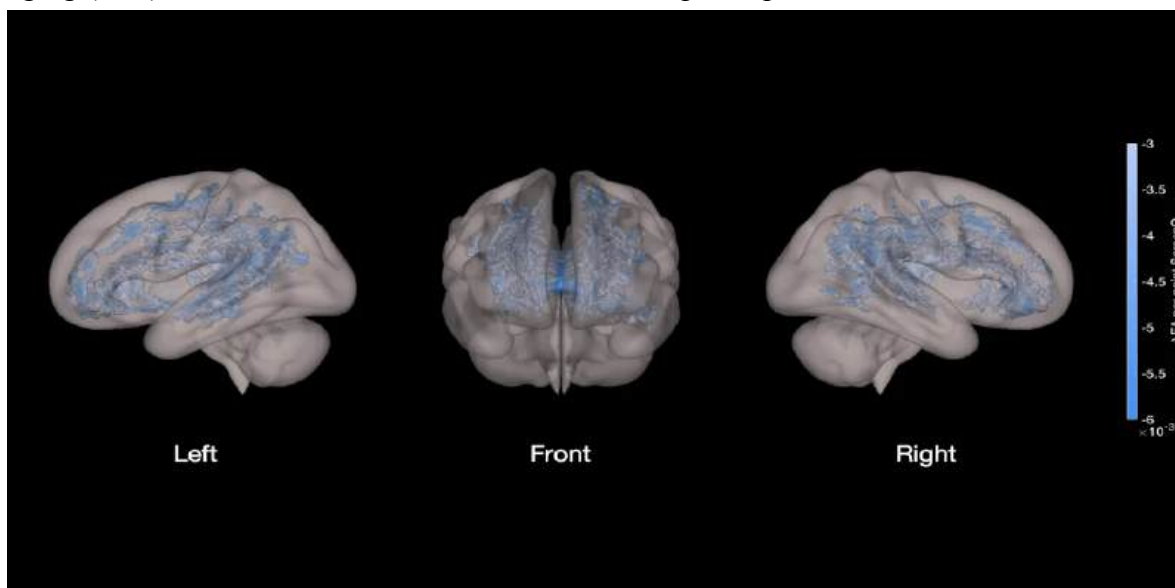


Here is a PET scan of a cocaine user. These images show exactly which part of the brain is affected by addiction, that is one's dopamine receptors. As described in the image, the scan on the left is the brain of the comparison subject, or the subject that is not a cocaine user. The other two scans show how gradually after the use of cocaine, the dopamine receptors become affected. Similarly, a 2012 Biotechnology study on those with Internet Addiction Disorder (IAD) found that "Disrupted dopamine function has been reported through a series of PET scans that consistently demonstrate at least a 20% decrease in D2 dopamine release and receptors in addicts' striatum" (Reduced Striatal Dopamine Transporters in People with Internet Addiction Disorder, NIH). This not only displays how both drugs and social media are highly addictive, but also shows how they have the same dangerous effect on the brain.

This effect is similar for other drugs that can cause addiction including methamphetamine, alcohol, cocaine, and heroin. And while social media and these drugs are starkly different, the fact that both these drugs, and social media affect one's dopamine levels and dopamine receptors, in the same addictive way, remains true.



However, these neurotransmitters aren't the only parts of the brain being affected, take into consideration the younger demographic of 3-5 year old pre-kindergarteners who have been surrounded by social media since birth. A 2019 JAMA pediatrics study, Associations Between Screen-Based Media Use and Brain White Matter Integrity in Preschool-Aged Children, found that the younger generation, that is 3-5 year olds in pre-school, exhibited slower neurological development, or more specifically, lower structural integrity of brain white matter within the tracts that support language and literacy skills. The study recruited a total of 69 participants through advertisement at a children's medical center who fit into the criteria: "children aged 3 to 5 years, born at at least 36 weeks' gestation, living in a household of native English speakers, without a history of neurodevelopmental disorder conferring risk of language delays, no previous or current kindergarten attendance, and no contraindications to magnetic resonance imaging (MRI), such as metal implants." It had each participant take standardized assessments of their literacy abilities through a vocabulary test (Expressive Vocabulary Test, Second Edition [EVT-2]), Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2), and a Get Ready to Read! (GRTR), as well as measured their screen times through a ScreenQ screen time assessment survey taken by their parents. Following this, the study utilized Diffusion Tensor Imaging (DTI) to observe the white matter within each participant's brains.



The study found that those who used screens more than the recommended one hour a day without parental involvement had lower development within the brain's white matter, as well as slower functioning of both their Wernicke's and Broca's area. Now, if this drastic of an effect is already displayed through an average screen time of 2.5-3.0 hours, one can only imagine the dangerous effects of screen time on the brains of teenagers who, on average, report using their phones for an upward average of 5 hours.

As per the National Institute of Mental Health, "ongoing changes in the brain, along with physical, emotional, and social changes, can make teens more likely to experience mental health problems. The fact that all these changes happen at one time may explain why many mental

illnesses—such as schizophrenia, anxiety, depression, bipolar disorder, and eating disorders—emerge during adolescence” (NIMH, 2023). Additionally, it has been found that too much or too little dopamine can cause “ADHD, depression, schizophrenia, a lack of impulse control, and a lack of motivation” (NIMH, 2023). Combine this with other associated pre-developmental issues and the newer social media algorithms that are designed to make users spend more time on screens- and the stark correlation between social media and addiction/mental ailment, becomes transparent.

As the famous Greek philosopher Democritus once said, “throw moderation to the winds, and the greatest pleasures bring the greatest pains.” An excessive amount of anything is dangerous, and while social media has its numerous benefits, like anything else it also has its setbacks. The teenage brain is still in development and until it becomes fully developed, it will continue to be more predisposed to any addictive behaviors or mental ailments. The key is moderation and awareness. Social media correlates to addiction and mental ailment but with moderation and caution “doom scrolling” and excessive social media use can be prevented.

Conclusion

Excessive social media usage has correlated to addiction and promoted addictive behaviors. With the way in which newly developed social media algorithms have specifically targeted one's dopamine receptor, there is a chance of the yet to be developed adolescent brain learning and adopting the same addictive behaviors and effect as that of drug usage. It has been found that excessive social media use not only negatively influences anxiety and depressive behaviors but also has a chance of promoting addiction to a younger generation, who in turn, can no longer process natural dopamine at the same level and whose baseline dopamine levels will skew dangerously.

Discussion

This connection between social media and addiction calls for immediate concern. While there has been a warning issued by the surgeon general, this issue calls for more awareness on the detrimental affects of social media upon long-time exposure at a young age. If this social issue continues to remain unsolved the younger generation, who already report being on their screens for an excessive amount of time may be predisposed to addictive behaviors. The fact that social media affects the same region within the brain as many dangerous drugs such as methamphetamine and cocaine displays just how easily it can cause addictive behaviors in the future. In fact, there may even be a chance that years down the road this addictive behavior of “doom-scrolling” becomes a gateway into further addiction. Regardless, it is essential that action be taken, especially in limiting screen time for teenagers and preventing social media use entirely for the younger generation of pre-adolescents. There is a call for attention towards this issue that must be addressed in an attempt to increase both protection and prevention.

Works Cited

- American Society of Addiction Medicine. (n.d.). Definition of addiction. American Society of Addiction Medicine. <https://www.asam.org/quality-care/definition-of-addiction>
- Gallup, Inc. (2023, August 17). Teens spend an average of 3 hours on social media per day. Gallup. <https://news.gallup.com/poll/512576/teens-spend-average-hours-social-media-per-day.aspx>
- Pew Research Center. (2023, December 11). Teens, social media and technology 2023. Pew Research Center. <https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/>
- U.S. Department of Health & Human Services. (2023). Protecting youth mental health: The U.S. Surgeon General's advisory. <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf>
- Harvard Health Publishing. (2023, June 6). Dopamine: The pathway to pleasure. Harvard Health. <https://www.health.harvard.edu/mind-and-mood/dopamine-the-pathway-to-pleasure>
- Science Photo Library. (n.d.). Healthy and addicted brains, PET scans - Stock image - C027/2374. Science Photo Library. <https://www.sciencephoto.com/media/1234567>
- Wang, Y., Li, X., Yang, L., & Zhang, J. (2023). The association between social media use and mental health among adolescents: A longitudinal study. *JAMA Pediatrics*, 177(8), 823-832. https://jamanetwork.com/journals/jamapediatrics/fullarticle/2754101?guestAccessKey=56c4b22b-ee5f-4594-bb23-c3813c9cccb1&utm_source=For_The_Media&utm_medium=referral&utm_campaign=ftm_links&utm_content=tfl&utm_term=110419
- Turner, B. (2011). The impact of social media on youth mental health: A review. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312312/>
- <https://www.mcleanhospital.org/essential/it-or-not-social-medias-affecting-your-mental-health>
- <https://www.center4research.org/social-media-affects-mental-health/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10129173/>

The Impact of AI on the Nursing Labor Market By Bill Wang

Abstract

Artificial intelligence is the latest in many waves of automation and seems poised to disrupt workers in many areas, including office professionals. In this paper, I narrow our focus to a single occupation: how will AI impact the nursing market? In this paper, I first review the history of automation, analyzing similar cases. Then, I discuss the potential impact of AI on the labor market in general, distinguishing among two dimensions: complementarity, which is when productivity is increased by incorporating AI, and exposure, which is how much AI will impact a certain profession. High complementarity and high exposure can grow the economy, leading to more jobs, but can also lead to increased wage inequality (Cazzaniga et al. 17). In low complementarity but high exposure models, AI technology can lead to substitution, which is when jobs can be replaced. In this model, income disparity between the 90th and 10th percentile decreased. However, in all models, capital income and inequality between the 99th and 90th percentile increased (Webb 4).

Finally, I look at nursing as a case study of the impact of AI on labor—focusing on recent technologies such as ChatGPT and LLMs, or near-future ones such as SORA, a text-to-video generator. Specifically, I observe the impacts on labor demand and how patients may react or feel about the usage of AI in nursing. Splitting nursing into a framework of separate tasks, I find that nursing is a high-exposure job with a somewhat even split between low-complementarity, medium-complementarity, and high-complementarity tasks. For future job prospects, I consider both saturated markets, like Los Angeles, with strong health care infrastructure, and unsaturated markets, like rural communities. In saturated markets, the increased productivity from recent AI technologies will likely lead to nurse layoffs, as there are few new customers that hospitals could attract. Even in unsaturated markets, a lack of digital infrastructure may disincentivize healthcare providers from expanding their services. However, an interview with an industry worker provides a more optimistic perspective. This paper concludes that the introduction of AI as an automotive technology into the highly exposed, medium-complementarity nursing occupation may find a decrease in employment in highly saturated markets, while desaturated markets may require a high entrance investment. In addition, I analyze AI as a transformative technology and find that it may cause task polarization between nurses, and employee reduction.

1 Introduction

The labor market refers to supply and demand, the model of the relationship between the quantity, want, and price, with labor as the focus commodity. Employers seeking to hire are the demand and workers who apply are the supply. The basic tenets of supply and demand apply. Factors that increase supply—increased competition for jobs from factors such as increased education, or immigration—with the same amount of demand, workers being hired, would decrease the price, or income. Decreased supply, like for blue-collar workers, and similar

demand, can increase price and income. This paper will focus on a factor that has the potential to drastically decrease the demand for new labor: AI technology.

According to a Health Workforce Analysis published by the Health Resources and Services Administration (HRSA) in 2022, there is currently a shortage of nurses within the industry (“Nurse Workforce Projections 2020-2035” 1). In 2024, the HRSA continued to predict a 10% shortage in RNs, and a 7% shortage in licensed practical nurses, but a surplus in nurse practitioners, anesthetists, and midwives in 2026 (HRSA, “Nurse Workforce Projections, 2021-2036” 2). The 2024 prediction of 2026 is a large increase from the 2022 prediction of 2025, with a 2% shortage in RNs, 5% in licensed practical nurses, with a similar surplus in other positions, so the shortage is seemingly growing larger (HRSA, “Nurse Workforce Projections 2020-2035” 2).

In 2024, nursing occupied 42% of all healthcare industry jobs (NU). As an indispensable laborer within the healthcare industry, it is important to recognize, predict, and react to disruptors such as AI technology. With accusations of disconnected leaders, a 91% increase in annual layoffs in 2023, and now with the threat of AI job takeover (with my later analysis showing that nursing is a high exposure and medium complementarity occupation), the importance of this analysis becomes augmented when placed within such a precarious industry (Lobdell, “December 2023 Challenger Report”).

It is flawed to analyze or predict the effect of highly speculative AI technologies, such as general AI, as arguments can be made for artificial intelligence to replace everything, so instead, AI technology will be limited to those currently accessible, such as ChatGPT, and near-future technologies, such as Sora, OpenAI’s text-to-video model (OpenAI, “Introducing Sora”)

I first begin with a historical background, exploring historical labor-saving technologies through a literature review, highlighting the First Industrial Revolution and the beneficial effects of increased productivity, and increasing life quality. I then highlight the accounting occupation, and how the advent of the mechanical calculator drastically changed the labor market landscape. Accounting experienced wage polarization, splitting into bookkeeping and accounting (Wootton and Kemmerer 93-94). Finally, I will briefly speak about some current AI technologies.

In Section 3, I discuss more modern innovations and find that AI technologies are slated to have a drastic effect on the labor market (Zarifhonarvar 14). I also discuss classifications created by an IMF paper: complementarity, when productivity is increased; substitution, when jobs can be replaced; and low exposure (Cazzaniga et al. 4). Historically, technological advancements have increased productivity, spurring economic growth, consumer demand, and jobs. Nevertheless, automation leads to job displacement. In low complementarity, high exposure occupations, income disparity between the 90th and 10th percentile may decrease. However, capital income and inequality between the 99th and 90th will increase (Webb 5).

Section 4 utilizes the International Standard of Classification of Occupations, which splits up certain jobs into necessary tasks. I analyze each task, using framework analysis, evidence, and upcoming technology to generally classify each task into low exposure, high complementarity, medium complementarity, and low complementarity. The threat of substitution is indicated by

low complementary tasks (Cazzaniga et al. 4). I find that when considering each task of equal value, nursing is a high-exposure occupation, with a medium level of complementarity.

Despite this, as I reason in AI in Section 5, higher productivity and complementarity can still lead to job disruption. Healthcare is considered highly saturated, and thus, mature companies may lose the incentive to expand using their higher productivity, increasing margins, and decreasing operating costs. This may lead to the unemployment of nurses. Furthermore, in unsaturated geographic demographics, a lack of digital infrastructure may mean that an initial expansion may be costly. Amongst already increasing layoffs, and decreased nurse satisfaction, I conclude that nursing may be in a precarious state, albeit not obsolete. Analyzing AI as a transformative technology also indicates task polarization and chance for substitution.

I conclude with excerpts from an interview with an industry worker. She points out a more optimistic point of view, believing that the social stigmas will check the expansion of AI technology, as well as several flaws that AI technologies may not accommodate. Her points are supported by evidence showing discomfort with AI in healthcare, and the potential for AI to solve the nursing staffing (and workload) crisis instead of increasing layoffs.

Overall, I find that nursing is a high-exposure, medium-complementarity occupation, and that in highly saturated markets, low chance for expansion may incentivize the reduction of employees, while less saturated markets may require a high entrance investment, and may require incentives other than profit. An industry worker points out that social stigma and the nurse staffing crisis may keep job displacement in check, however. In addition, when analyzing AI as a transformative technology, I find that AI may lead to task polarization, but also risk job displacement.

2 Historical background

Automation encounters heavy scrutiny and conversation, due to the capabilities to replace many workers undertaking more routine jobs, but also enhance productivity, giving workers more resources such as mapping software for taxi drivers and perhaps even creating more jobs than it takes (Webb 1-5). As we enter a new era of technological innovation, it becomes incumbent to analyze and prepare for new technologies, to regulate, or expedite the creation of such technologies, and to make sure that we all benefit from them. Artificial intelligence seems to hold the power to automate many tasks previously thought immune to automation, sparking understandable concern for the livelihoods of many. AI is an example of an LST or Labor Saving Technology. Although AI technology and its potential—whether as a boon or bane—seems to be an unprecedented dilemma, I can look to history for help. Previous reactions and outcomes reflect themselves over time, allowing us to learn from the successes and failures of the past.

The First Industrial Revolution, beginning around the mid-18th century, was delineated by the mass mechanization of low-skill and manual jobs. Before the mechanization of certain processes, production occurred in a decentralized cottage industry, rather than the centralized factory systems of today. This allowed for the production of goods at home, which would then be sold to business owners. With the advent of innovations such as the spinning jenny (a 1764

invention that allowed a single machine to spin multiple threads at a time), the cotton gin, and the steam engine, the cottage industry lost its position as the primary source of production. Similar advancements within the agricultural sector led to a decrease in demand for hard labor. The resulting unemployment drew rural workers to cities, spurring urbanization. Such workers often faced harsh working conditions, due to a lack of regulation, and meager wages (Floud and McCloskey 17, 19, 24, 144, 341, 371).

However, at the same time as the supply of goods increased, the price subsequently decreased. While machines replaced human labor, this allowed for the standard of living to increase for the common worker, boosting productivity, as well as decreasing the time and labor required for each product. Mass production increased the capabilities to produce what is considered modern necessities, such as ready-made clothing (Floud and McCloskey 12, 385).

More modern LSTs include electricity, robots, and computers, each bringing a wave of discourse and seemingly higher stakes than ever. In 1958, America encountered a severe economic downturn, dubbed by the magazine *The Nation* as the “automation depression.” The magazine argue that even though more and more Americans were looking for jobs, job demand continued to remain stagnant, if not reduced due to the onset of “worker-displacing innovations.” Others saw it as a part of the natural business cycles, from businesses acting to maximize profits. Corporate executives argue that consumer demand from expanded businesses would create more jobs to compensate (Wartzman). Yale Brozen, University of Chicago economist, argues that while 13 million jobs were removed during the 1950’s, 20 million more jobs were created (Brozen 16). Rick Wartzman, in an article detailing the “automation depression,” argues that while more jobs could have been produced, the displacement led to the destruction of certain sectors, causing unemployment and displacement as workers were forced to find new jobs, or re-educate themselves, something that the U.S. has historically struggled with accommodating (Wartzman). But whereas mechanization, robots, computers, and other past LSTs had a larger effect on simple jobs such as manual labor, factory production lines, and cashiers, AI puts high-level jobs at risk, and could thus induce different effects.

A specific example can demonstrate such effects. Accounting is often considered to be a high-level job, even more so in the 19th century. Initially starting with handwritten records in journals, the profession was drastically changed by technological developments such as typewriters and basic calculators. Such devices allowed for the rapid calculations that facilitated such data processing. The advancement of accounting technology also created a new industry, with companies emerging such as IBM, which initially focused on early data processing systems (Wootton and Kemmerer 92, 105).

However, mechanical accounting technologies (such as the mechanical calculator) seem to have a polarizing effect on the tasks taken by its workers, changing the composition of the two branches of bookkeeping and accounting. Bookkeeping was primarily handled by women, due to perceived dexterity, and substantially lower wages, and mainly concerned itself with storing and organizing transactional records (Wootton and Kemmerer 94). Meanwhile, men largely got the more professional and high-skilled job of accounting, which included the analysis of such data.

Ultimately, however, the introduction of accounting machines allowed banks to cut the number of employees (Wootton and Kemmerer 93, 115). Here is a polarizing effect on the workforce, with one less-educated group being delegated to more menial tasks that were not yet automated, and facing wage stagnation, while the other was able to take on even higher-levelled jobs and benefit from the increased productivity of accounting technologies (Wootton and Kemmerer 5). This case study is later applied to nursing, utilizing the historical effects to predict the transformative effect AI technology may have.

Today, according to a 2021 EMSI labor analysis, bookkeeping has a median salary of \$42,411, while accounting has a median salary of \$73,750. When looking at the top ten percent salaries, accounting has a projected income of over two times that of bookkeeping, demonstrating higher initial salaries, and higher growth opportunities. Furthermore, accounting demand continues to grow, with a projected 6.1% increase over ten years, while bookkeeping is projected to decrease by 1.9% over the next ten years (Franklin University). So, those who were able to adapt to the new technologies benefited greatly, while others were left behind.

The phenomenon of AI technology has taken the economy by storm, with companies struggling to come out on top of this new wave. Innovations include Devin, an “AI software developer” whose demo embroiled the online development community in discussion, Sora, a powerful generative AI model capable of creating high-quality videos, and ChatGPT (Cognition Labs, “Introducing SORA”). I now explore how this wave of AI has impacted the labor market.

3 The Effect of AI on the Whole Labor Market

An LLM, or large language model, uses machine learning to understand and generate human languages. In the 1960’s, MIT researcher Joseph Weizenbaum developed the first chatbot, ELIZA, which processed natural language by identifying high-priority keywords, and then following “rules” associated with such keywords. But while ELIZA could hold conversations, it tended to speak in platitudes and verbiage. Weizenbach noticed that during a therapy session, a therapist often did not add much to a conversation, instead building off the patient, who would be the one carrying the conversation. For example, if a patient were to mention his father, ELIZA would respond with something along the lines of “Tell me about your father” (Weizenbaum 36-7, 42). By restricting ELIZA to a therapeutic function, Weizenbaum created a chatbot so potent that it spawned the term “Eliza Effect,” which is when humans project real human traits onto computer programs. This phenomenon points to the idea that LLMs can substitute certain traits thought irreplaceable, such as emotional welfare from a “human presence” which may be relevant to the nursing profession. The National Eating Disorder Association dissolved their eating disorder hotline, mass firing staff workers—a daring move staff claimed was triggered by their attempt to unionize—in favor of an AI-powered chatbot. If the replacement was indeed triggered by a unionization attempt, this may set a concerning precedent (Barr). This effect is later mentioned in my analysis of nursing tasks, where the nursing occupation is identified to be high-exposure, including in certain aspects previously thought more resistant to AI.

Later advancements, such as the LSTM (Long Short-Term Memory) and word embedding further spurred increasingly larger and larger models (Toloka). In 2019, Google researchers introduced BERT, a 340 million parameter model that understood words in context to a larger sentence, rather than individually. The 2019 BERT update to Google greatly improved search queries—the suggestions that pop up when you begin typing in a search bar—and search results. This is especially true of queries that use propositions like “to” or “for” (Nayak).

“Here’s a search for “2019 brazil traveler to usa need a visa.” The word “to” and its relationship to the other words in the query are particularly important to understanding the meaning. It’s about a Brazilian traveling to the U.S., and not the other way around. Previously, our algorithms wouldn’t understand the importance of this connection, and we returned results about U.S. citizens traveling to Brazil. With BERT, Search is able to grasp this nuance and know that the very common word “to” actually matters a lot here, and we can provide a much more relevant result for this query” (Nayak).

So, these LSTM models enabled a much more complete understanding of text than ELIZA’s keyword approach.

The release of ChatGPT 3 (Generative Pre-Trained Transformer) in November 2022 gained mass attention due to its powerful text generation abilities and ease of use. ChatGPT 4 can analyze and comment on visual data as well (OpenAI, “Say Hello to ChatGPT-4o”). These models can help with various practical problems, from drafting cover letters and emails to programming coding problems. Thus, they represent a huge step up from earlier models, which often could only help in specialized domains.

As AI technology continues to grip public attention, businesses have begun to take advantage of such high-technology tools to come out on top. Companies and countries have entered a new technological race. OpenAI, in February 2024, gave a demo for SORA, an AI tool that generated photorealistic, high-quality videos (OpenAI, “Introducing Sora”). In March 2024, Cognition Labs also released a video demonstration for Devin, an AI-powered software engineer capable of taking prompts, fixing bugs, and learning (Cognition Labs). Due to the drastic effect that ChatGPT and AI technologies can have on the automation of jobs, the impact of such technologies has been deeply scrutinized.

For example, one study analyzed the International Classification of Occupations (ISCO), which organizes and contains job information. They used text-mining for certain keywords that indicate exposure to AI, finding that 32.8% of occupations could be fully impacted, 36.5% might experience a partial impact, and 30.7% were likely to remain unaffected (Zarifhonarvar 14). A fully impacted occupation indicated the possibility for total replacement, bringing cost efficiency and general efficiency to employers at the cost of job loss. A partially impacted occupation was more likely to be assisted by technology, making more free time or leading to more productive work at the risk of lesser job loss and wage stagnation, while an occupation with no impact usually required specific specialization or physical labor.

Some specific jobs at risk, as outlined by the text-mining study, are computer science, media management and creation, lawyers, customer service, and market analysts. As such, high-level jobs are the ones currently at risk from AI technology, as compared to previous LSTs which mostly impacted low-level jobs (Zarifhonorvar 4-5).

The rise of AI is similar to a new industrial revolution, especially since AI can affect global markets. Since AI affects more high-skill level jobs, advanced economies, which hold higher education rates, are more at risk from the impact of AI. At the same time, more advanced digital infrastructure enables advanced economics to better leverage the potential benefits of AI technology (Cazzaniga et al. 11).

AI, as an automotive technology, has the potential to increase output, which may stimulate more labor demand and overall labor income through the opportunity for expansion, and economic growth. However, this could exacerbate existing inequalities. For example, women are generally employed in more AI-exposed occupations, and older citizens may struggle to adapt to new technologies and find new employment. There may be obstacles such as societal resistance to AI technologies due to ethical reasons, as well as modern technological divides between those who can own, and those who do not own AI technology. Displacement of jobs may also lead to downward pressure, in which a larger number of unemployed people leads to lower income from competition among peers (Cazzaniga et al. 5).

At the same time, AI technology can help reduce inequality within countries, as the government can utilize resources to increase assistance and productivity of social services. A study by the IMF finds that digitalization has improved the effectiveness of social services, and thus the reduction of inequality. For example, LLMs could perhaps be trained to give financial advice to users, utilize public services, or be trained on medical data, to reduce medical bills from hospital visits (Amaglobeli et al. 13, 18-9).

Another paper from the IMF continues to investigate the effects of AI technology on the labor market by expanding the concept of AI exposure into three sections: high exposure and high complementarity; high exposure and low complementarity; and generally low exposure (Cazzaniga et al. 4). High exposure and high complementarity jobs include jobs such as judges, who can gain from the textual analysis offered by LLM technologies, but would be able to resist displacement due to social barriers, instead having their productivity “complemented” by AI technologies. A high-exposure and low-complementarity job would be a telemarketer, or customer support, in which direct human contact is usually not needed, and generation of text, or voice can suffice (Cazzaniga et al. 10). In a later analysis of the nursing occupation, nursing is identified to have a high exposure, due to the vulnerability of certain administrative tasks, and has a medium complementarity.

The IMF paper argues that if AI has low complementarity with labor, labor income inequality will decline. This is because several high-paying jobs would be eliminated, as AI could entirely replace some computer scientists, or lawyers. This displacement effect would cause income inequality to lower. However, if AI is highly complementary to labor, then jobs with high exposure like computer scientists and lawyers would be much more productive and

could demand higher wages. Lower-income jobs tend to be less likely to be exposed to AI, and thus would not benefit from the complementarity. Thus, income equality would increase. Unlike labor income inequality, the paper found that in all scenarios, capital income and wealth inequality increased (Cazzaniga et al. 4). While these papers focus on the impact of AI on the entire labor market, for the rest of this paper I narrow my focus to just one job: nursing.

4 The Effect of AI on the Nursing Occupation

Medical care, directly tied to human lives, is a significant industry. It not only competes with other providers but also against death, necessitating constant innovation to combat disease. Thus, AI technology will be implemented, will begin to be implemented, and has already been partially implemented in the medical care industry. Health IT Analytics predicts that by 2025, global AI healthcare spending will reach \$36 billion. China aims to become a global leader in AI technology by 2030, and in 2019, the US passed an Executive Order on Maintaining American Leadership in Artificial Intelligence (Bresnick, Robert). In this section, we'll analyze how recent and near-future AI technology might impact the work that nurses do daily.

The International Standard of Classification of Occupations (ISCO) is a database of occupational data, commonly used for research regarding occupations, and skills. ISCO breaks down each job into a series of necessary tasks, which I individually analyze for the specific effects that recent and near-future AI technology can have—complementing, or replacing (classifications an aforementioned IMF paper discusses)—upon the nursing occupation. The following is directly taken from ISCO:

- (a) planning, providing and evaluating nursing care for patients according to the practice and standards of modern nursing;
- (b) coordinating the care of patients in consultation with other health professionals and members of health teams;
- (c) developing and implementing care plans for the biological, social and psychological treatment of patients in collaboration with other health professionals;
- (d) planning and providing personal care, treatments and therapies including administering medications, and monitoring responses to treatment or care plan;
- (e) cleaning wounds and applying surgical dressings and bandages
- (f) monitoring pain and discomfort experienced by patients and alleviating pain using a variety of therapies, including the use of painkilling drugs;
- (g) planning and participating in health education programmes, health promotion and nurse education activities in clinical and community settings;
- (h) answering questions from patients and families and providing information about prevention of ill-health, treatment and care;
- (i) supervising and coordinating the work of other nursing, health and personal care workers;

(j) conducting research on nursing practices and procedures and disseminating findings such as through scientific papers and reports (International Labour Office).

By splitting the occupation into a series of tasks, and evaluating each task on its complementarity, or lack thereof, I can better analyze how AI impacts the field of nursing, allowing us to determine the exposure and complementarity of their intersection. The concept of general AI, or more speculative technologies, will likely not be discussed as its theoretical functions would be able to exceed common limitations. A downfall of this analysis is that it treats each task as equal when in reality, some responsibilities are more labor and time-intensive than others. This caveat is further addressed in the Section 5. The following is an interpretation of each task, and its complementarity (or lack thereof) for now and in the future.

A) Task A is to “[plan], [provide] and [evaluate] nursing care for patients according to the practice and standards of modern nursing” (International Labour Office). I argue that AI is capable of doing this, and that certain aspects of nursing, such as “human presence” are more substitutable than thought, presenting a moral hazard. The standards of modern nursing are later explored within our analysis of task A.

Planning and evaluating nursing care can be effectively handled by generative AI due to its analytical capabilities and access to abundant data (with confidentiality being a potential hurdle). However, the actual provision of care is more complex and not as easily substitutable.

Provision 1.3 of the American Nurses Association Code of Ethics states that “optimal nursing care enables the patient to live with as much physical, emotional, social, and religious or spiritual well-being as possible, and reflects the patient’s values. Supportive care is particularly important at the end of life to prevent and alleviate the cascade of symptoms and suffering that are commonly associated with dying” (American Nurses Association, “Code of Ethics” 1-2).

New technological developments tend to offer more development for the physical care of patients, but AI technology, especially generative AI, and the more speculative emotional AI (although due to the ELIZA effect, the necessity of emotional AI can be contested) can further the emotional care of patients via technology. While healthcare often requires a human touch, modern generative AI has facilitated the development of emotionally responsive robots, also known as social or companion robots.

One example is the aforementioned ELIZA, the therapist-robot, which convinced its users to project human emotions onto it, becoming attached, and creating the ELIZA effect (Weizenbaum 42). Another, more recent example, Sophia, a social robot initially created to accompany seniors was personified to the extent of earning Saudi Arabian citizenship (British Council). Even more recent are the swarms of new chatbot programs spawned through the emergence of ChatGPT in 2022, such as character.ai, a program that allows one to talk to a character from a popular piece of media, and even the aptly named AI Chat Girlfriend.

These technologies seem slated for even more future advancements—and soon. ChatGPT-4o, receiving the grandiose appellation of “Omni” for its ability to handle audio and

visual mediums was released in May 2024. Some notable features shown in its live demos included a voice deep fake audio call using preset voices, and computer vision capabilities (OpenAI, “Say Hello to ChatGPT-4o”). Increased intractability and more “human” features can augment ELIZA effect and perceived empathy. With these recent advancements in mind, the usage of generative AI in health services seems more conceivable.

A qualitative study explores various need-based communications within nursing and classifies informal education and the inducement of calmness in the patient as categories of communication (Fakhr-Movahedi et al.). The study separates these responsibilities into categories, which are listed below. By individually analyzing each responsibility, I can systematically analyze the extent to which an AI tool will be effective, and thus, the complementarity of task A.

Informal education of the patient:

- i) Explanation about medication,
- ii) Medical and nursing interventions
- iii) Patients' preparation for diagnostic procedures
- iv) Discharge preparation

Inducing calmness to the patient:

- i) Respect to patients
- ii) Nurses' patience to patients' requests
- iii) Sympathy with patients
- iv) Listening to patients' concerns

Obtainment of trust

- i) Understanding the patients' conditions
- ii) Accepting the patients (Fakhr-Movahedi et al.).

Analyzing each responsibility that Fakhr-Movahedi et al. outlines, we find that AI is effective at fulfilling each responsibility. Regarding informal education, generative AI is rather adept at communicating formal information, and given accurate information and diagnosis would be capable of dealing with less emotive aspects. When it comes to inducing calmness, once again, the ELIZA effect can be brought up, demonstrating the ease at which patients or users can be deceived into the projection of therapist-like behavior onto computer programs (Weizenbaum 42). Respect, patience, and sympathy are traits that can easily be implemented through the application of rhetorical choices and techniques. However, the ELIZA effect operates on the basis that the patient does not know they are communicating with a robot. The PEW research center finds that many surveyees are concerned with the aspect of humanity in healthcare and that 60% are uncomfortable with AI in healthcare (Tyson).

Finally, the obtainment of trust can be encoded within the rhetorical choice to “trick” recipients into believing in the emotive senses of generative AI. This could be supplemented by computer vision or tone recognition technology. With breakthroughs in AI technology, affective

computing, which deals with the interpretation, processing, or even emotion simulation models may be a keystone to creating more “human” communication. AI is capable of communicating diagnosis information “sensitively” by tricking users into projecting real emotions.

Advancements in audio deep fakes (and maybe even facial) could easily supplement the convincing power of generative AI in convincing the patient of a model's humanity. Automating emotional and spiritual well-being in healthcare presents potential moral hazards. There are two primary approaches: fully convincing patients that they are interacting with a human, or being transparent about the AI's role, each with its ethical implications. A company can either choose to completely trick patients into believing that the model is a real person, potentially receiving benefits in customer satisfaction, or let patients know that they are receiving a diagnosis from an AI and incurring the wrath of the 60% uncomfortable with AI in healthcare, and hoping that emotions are projected (Tyson). From a regulatory and moral standpoint, the second strategy may remain more feasible. From a profit and service quality standpoint, when considering opportunity cost, the second strategy may also remain feasible.

Public attitudes towards AI may become more positive. A female Snapchat AI salaciously earned \$72,000 in its first month, and Late Checkout CEO Greg Isenberg predicted the AI girlfriend market to reach one billion dollars (Paris, Zilber). Nevertheless, reported acceptance of AI in affective computing/well-being in AI girlfriends may not be indicative of acceptance within healthcare due to drastic differences in market demographics and expectations of professionalism. Without cultural influences, I find that with current, not near-future or speculative technologies, the emotional aspect is, contrary to popular belief, substitutable. This may decrease labor demand as increased productivity from this one task alone may justify the nurse layoffs, depending on the choice of whether to expand or not, as is explored later.

Other methods of providing nursing care such as physical care, will likely not be substitutable with current or near-future AI technology due to nuance and diverse factors, as well as adaptability, and instead complement efficiency at most. Certain interventions, like bedside care, are labor intensive, and a physical presence is unlikely to be a replaceable commodity.

B), C) Tasks B and C require coordination of “the care of patients...with other health professionals and members of health teams,” and in “developing and implementing care plans for the biological, social and psychological treatment of patients” (International Labour Office). I argue that AI is capable of complementing this strongly. Due to the tasks’ similarity, and shared requirement of coordination, they are analyzed together.

As it uses machine learning, and data from the past, AI technology should be exceptional when it comes to processing and utilizing data. Nevertheless, coordination of the care of patients does not necessarily need the usage of such high-technology applications.

However, it is possible for AI to completely automate schedules, meaning that what human input previously required would become completely redundant, and AI technologies can substitute this task. As of right now, generative AI technology is not fully trustworthy, and would

likely be complementary, needing human confirmation. However, in the future, this task may be fully substituted, with each healthcare provider being told exactly what to do.

Generative AI will likely be particularly adept at creating plans due to the protocol-based nature and ease of obtaining data, although as of right now, implementing them physically, as previously mentioned, would be a challenge. As of now, plans would only complement healthcare workers. As of later, they could potentially do all the administration.

Nevertheless, concerns about bias and data privacy arise. Improper datasets with biases, as is commonly mentioned, can become exacerbated through repeated reinforcement and exacerbate racial divides and inequality. Meanwhile, with large datasets required, data privacy may be a concern for patients as well as healthcare providers (Pailaha). Furthermore, as is explored in Section five, nurses commonly partake in fast-paced situations in which high adaptability is needed. Current and near-future AI technology is not highly adaptable in such scenarios. In addition, in terms of communication, the usage of AI may facilitate longer-paced communication, such as scheduling events. But again, in fast-paced situations in which high reaction speed is required, direct communication is favored over indirect. Thus, I find that this task is somewhat substitutable, with the nuance of fast-paced situations.

D) Task D requires “planning and providing personal care, treatments and therapies” (International Labour Office). Section A previously discussed the implementation of mental health support using AI technologies, particularly generative AI. Again, more physical implementations would be difficult to carry out. While AI technology holds massive potential when it comes to more systematic, narrow decisions like monitoring responses to care, it is unlikely to fully replace this task.

E) Task E is more physical, requiring “cleaning wounds and applying surgical dressings and bandages” (International Labour Office). I argue that task E is low exposure due to inadequate technology.

As of now, AI is unlikely to substitute nurses, especially when it comes to more physical scenarios and scenarios in which a nurse may need to be adaptable. Computer vision technology is already being implemented into medical care, such as in X-rays, but is only used within the realm of complementarity (Gao et al. 3-4). The pairing of robotics and artificial intelligence, however, is becoming a large realm of interest, and will likely encounter significant growth.

As data collection for an AI trained on cleaning wounds should be easily and cheaply obtained, cleaning wounds and applying dressings is plausible, and may lead to substitution. However, because there are so many possible wounds, it is difficult to say that an AI without a high degree of adaptability would be able to effectively apply dressings and bandages. Because current and near-future AI technology does not seem to be poised to complete highly adaptable, hands-on work, not complementing nor substituting, there is no effect.

F) Task F refers to “monitoring pain and discomfort” and “alleviating pain using a variety of therapies, including the use of painkilling drugs” (International Labour Office). I argue that while it may seem to be a systemic task, with computer vision, and possibly there may be caveats. Thus, current and near-future technologies will only be highly complementary, not substitutable. In the task F analysis, I explore the capabilities of AI to monitor pain, and a moral hazard that may arise from the failure to do so effectively.

Monitoring pain and discomfort, and then alleviating said discomfort seems to be a rather systematic, and thus, replaceable task. Nevertheless, societal resistance may inhibit AI technology from actually doing so. Furthermore, a certain ethical concern arises from the abuse of pain-relief medications: a patient may be able to trick AI into recommending pain medications, whereas a doctor or nurse may be better at recognizing body language or seeing the same patterns with a particular patient. The same goes the other way. In a later-described interview with a Nursing Assistant, the assistant brings up the example of chronic pain. A patient with chronic pain may become desensitized to the pain such that they may not be so physically affected. AI facial recognition technology and computer vision may fail to recognize this pain, and thus fail to properly prescribe medicine.

G) Task G refers to “planning and participating in health education programs, health promotion, and nurse education activities in clinical and community settings” (International Labour Office). Generative AI technology is already being used in education (HMH). Through chatbots and adaptive learning, AI can not only complement but even substitute teachers. On even a more conservative stand, AI can at least expand access to education.

H) Task H refers to “answering questions from patients and families and providing information about prevention of ill-health, treatment and care” (International Labour Office). The communication skills of AI technology are explored within section A. However, the feasibility of Task H is further explored below, in which I argue that AI technology is capable of effectively communicating, and is already being used, but has certain caveats.

The World Health Organization (WHO) has developed a Smart AI Resource Assistant for Health, or SARA, which utilizes OpenAI and is trained on a medical dataset (Brogan). Since ChatGPT and generative AI are rather open to the public, generative AI can increase access to medical care, although an AI is unlikely to be considered more trustworthy than a medical professional, and certain prescriptions may lead to adverse side effects.

Although there are caveats, AI is far more capable of answering questions with current technology, being able to take on more workload than an actual expert, even if expert opinions are considered more valuable. Nevertheless, due to the massive difference in productivity, and even if expert opinions won't become obsolete, this is highly substitutable. In the future, as AI becomes more normalized, and advanced, generative AI can substitute medical professionals, and perhaps even exceed them.

I) Task I is “supervising and coordinating the work of other nursing, health and personal care workers,” which overlaps with B and C in terms of coordination (International Labour Office).

J) Task J refers to “researching nursing practices and procedures and disseminating findings such as through scientific papers and reports” (International Labour Office). I argue that AI is already being used in research, although a highly complementary task, not substitutable.

As of now, AI requires a dataset and is not highly adaptable. Thus, when it comes to research, AI is incapable of fully replacing researchers. However, generative AI can play and already plays a large role in complementing researchers. For example, generative AI and LLMs are capable of consolidating literature and presenting it in a more simplified view. In addition, they can be used to turn qualitative data into quantitative, such as the usage of LLMs in data mining, as well as generally evaluating qualitative information at a rapid pace. The complementarity of AI technology will likely only increase as it becomes more normalized.

The following task matrix visualizes the results of my findings to better clarify the level of exposure and complementarity of nursing. The task matrix demonstrates a high degree of exposure, with around a medium level of complementarity overall.

Task Matrix

	Exposure (L/H)	Low Complementarity	Medium Complementarity	High Complementarity	Comments
A	H		✓		Planning, but not providing is substitutable.
B	H		✓*		Varies.
C	H	✓			
D	H		✓		
E	L				Low Exposure
F	H			✓*	AI may fail to capture nuance.
G	H			✓	
H	H	✓*			Planning, but not providing is substitutable.
I	H		✓*		Varies by situation.
J	H			✓	

5 The Effect of AI on the Nursing Labor Market

In Section 1, through the literature review, I found that during the First Industrial Revolution, the creation of the steam engine, among various other innovations allowed for the creation of the mass-production industry and factories. However, as technology continued to advance, concerns about the loss of jobs from LSTs sparked civil disobedience. For example, in 19th-century Britain, the Luddite movement was sparked by power looms, leading to protests in which new machinery was destroyed. Ultimately, however, the Luddites were unsuccessful, and the machinery was still used (The National Archives). While the First Industrial Revolution did indeed displace many jobs, the absolute standard of living and employment opportunities still increased over time (Floud and McCloskey 385).

In Section 2, I discussed the various effects that Artificial Intelligence technologies would have on the market as a whole. As of now, it is difficult to definitively determine the exact manner in which the job market for nurses will be affected.

In Section 4, I also determined that nursing is a high exposure occupation, with medium levels of complementarity. This information is further used in this analysis.

More likely than not, AI technology will be too powerful to be unused in the medical care industry. Nurses will likely need re-education or training to utilize AI technologies. But some nurses are resistant to the idea. In late April/early May, 2024, the slogan “Trust Nurses, not AI” arose, in which protesting nurses and healthcare workers argue that AI technologies would lead to more problematic standards and practices within the medical care industry. For example, they argue that AI was more susceptible to biases and inaccuracies, cited data privacy violations, and could lower the skills of workers who grow too reliant on them. They also argue that AI is often released in an unfinished state, being pushed out early for quick profits and increased margins over actual concerns about healthcare accessibility or quality (Walker).

Yet despite growing concerns about modern AI technologies and their effect on displacement, industry trends seem already geared towards workplace displacement. The AACN finds that currently, there is a shortage of nurses within the industry, predicted to reach a shortage of 78,610 full-time RNs in 2025 and a shortage of 63,720 full-time RNs in 2030 (“Nurse Workforce Projections 2020-2035” 1). Furthermore, even after the stressors of COVID-19, the possibility of shortage continues to increase, in 2021, two-thirds of nurses would have recommended their job, but only half did two years later (Rosseter).

The MIT Sloan points towards disconnected leaders, who have decreased their demand for new nurses within the occupation, and choosing to increase patient loads for no extra compensation (Sull and Sull). If leaders continue to thrift and prioritize saving money over quality of care and expansion, nurses will certainly experience a decrease in job demand, with or without the supplemental driving factor of modern (and prospective) AI technologies. According to a study by Challenger, Gray, and Christmas, the healthcare sector cut 58,560 jobs in 2023, a 91% increase from the previous year (Lobdell, “December 2023 Challenger Report”). The evidence for disconnected leaders is correlated with decreasing nursing job satisfaction, with the aforementioned decreasing job recommendation willingness.

My Section 3 analysis on the various nurse tasks, and their level of complementarity versus substitution finds that from the perspective of current and (estimated) near-future AI technologies, it is unlikely for the demand of the nursing occupation to become obsolete. At the same time, I find that certain tasks believed to be unable to become obsolete, such as the preservation of emotional well-being, are more at risk than thought. It becomes unnecessary to analyze the impact of AI technology from an extremely prospective and speculative framework as AI technology at an advanced enough level can be argued to replace humans.

Yet while nurses are unlikely to become obsolete, the labor market for their occupation could still increase or decrease. For example, in a saturated region such as California in which 92% of Californians already have health coverage, as the Legislative Analyst's Office finds, healthcare providers are unlikely to find more customers, and thus with the additive productivity from AI technologies, choose to lay off nurses to meet the demand of nurse work done (Petek 2).

In Section 4, I used an organized structure to analyze the impact of AI technology on nursing by splitting the occupation into a list of tasks and then evaluating each task. From this point of view, the occupation of nursing has a widespread level of complementarity, skewed slightly towards high complementarity. However, this analysis fails to differentiate between tasks, depicting them as equally labor-intensive when they are not. Furthermore, as I argue earlier even if simply complementary, jobs may still be lost.

In Section 3, I found that more administrative and formal tasks, such as communicating various information, scheduling appointments, and creating and coordinating various interventions have a lower degree of complementarity, while direct, physical care and fast-paced situations have either low exposure or high complementarity.

However, a 2020 report finds that nurses only spend about 21% of their time on direct patient care (Sun and Cato 27). This means that the task with the lowest exposure to AI takes up a small fraction of a nurse's time, leaving nurses vulnerable to AI exposure on the tasks that take up most of their time, and suggesting lower complementarity than implied in the task matrix.

Unsaturated markets may face other issues. In a study published in the Mayo Clinic Proceedings: Digital Health, it was found that healthcare coverage was correlated to internet access, with countries with the lowest broadband access (64%) being accompanied by the highest healthcare resource constraints and accessibility (Cuadros et al.). AI technologies require significant digital infrastructure, such as reliable power sources, and high-speed internet. However, because less-saturated markets may lack the necessary digital infrastructure, healthcare providers will be less able to deploy AI solutions.

On the surface, this may seem positive for nurses, who may expect increased labor demand in these unsaturated markets. However, there are two reasons that this may not be the case. First, AI will likely lead to increased profit for companies (as well as improvements to life quality). Since healthcare companies can expect higher profit margins for AI, they will be disincentivized from expanding to unsaturated markets, where they would have higher costs and lower profits. Second, nurses may be less willing to relocate to these areas. For example, Challenger, Gray, and Christmas find that due to increasing real-estate costs and the COVID-19

pandemic, job seekers are increasingly less willing to relocate to a job, even less so to an area with lower digital infrastructure (Lobdell, “Move to a New City for Work? No Thanks”). These will also contribute to a lack of willingness to expand services for healthcare providers.

Section 6 provides a different perspective on unsaturated markets, showing that from both a nonprofit and profit-driven frame of mind, AI technologies may provide an incentive for expansion—and therefore labor market demand increases. With layoffs already occurring before widespread AI implementation, and decreasing labor demand in saturated markets without corresponding increases in unsaturated markets, the job prospects for nursing appear challenging.

As a disruptive technology and an LST, however, AI technology has a larger effect than just increasing or decreasing jobs created. I may be able to expect an evolutionary effect on the job of nursing, and changes in the traditional tasks and responsibilities a nurse will undertake, having a more transformative effect on the labor market.

In a brief case study within my literature review, I covered the evolution of accounting due to the mechanical calculator. Generally, a mechanical calculator would have a highly complementary effect on accounting, being able to substitute various calculation tasks an accountant might make while complementing the accountant's job by increasing productivity. I also predict nursing to have a high degree of complementarity, substituting, or nearly substituting various tasks such as developing nurse intervention plans. Another similarity lies in that both nursing and accounting (at the time) were considered high-level jobs that were affected by an LST that targeted high-level jobs: the mechanical calculator then, and various AI technologies now (Wootton and Kemmerer 109, 114).

The accounting occupation undertook a transformative effect, splitting into the separate roles of bookkeeping, a lower-level occupation, and modern accounting, taking on more roles such as analysis of the actual data. Banks found that they could employ less accountants, while wage, task, and education levels between bookkeeping and current accounting continue to become more polarized. Given the similarities, I could also mirror the effects of accounting on nursing (Wootton and Kemmerer 93, 115).

Albeit speculative, a polarizing effect on nurses may manifest in “physical care nurses,” and “overseer nurses.” Physical care nurses would deal with the low exposure tasks that require a human presence such as feeding assistance, medicine administration, and applying bandages. While seeming like a “lower-level” occupation, this might not be the case. Bookkeepers had higher exposure to calculators, while here, “physical care nurses” will not be exposed to AI technology. Thus, their education level and wages will remain somewhat similar to “overseer nurses.” Being highly specialized individuals, “overseer nurses” would be the ones to quickly interpret and analyze AI tasks, usage, and results, rapidly giving the sign-off on care plans and schedules. Similar to how accountants had to pick up data analysis, “overseer nurses” may also require technical expertise to operate AI technology. I could also predict a polarizing effect in “physical care nurses,” and “overseer nurses” due to differing levels of specialization.

Such a transformative effect would lead to job displacement, and reeducation. An ethical concern arises from older nurses who may not be able to attain reeducation as easily as younger

ones, and various inequities propagated by the digital divide. Concerningly, regardless of whether AI is less or more transformative, both perspectives seem to allow a decrease in the number of nurses, as labor-saving technologies allowed businesses to employ fewer accountants.

Thus, I predict that in saturated markets, introduction of AI may replace workers, while unsaturated markets require high entrance investments, and AI as a transformative technology will cause task polarization among nurses, while still allowing for employment reduction.

6 The Effect of AI on the Nursing Labor Market: An Industry Worker's Perspective

However, an interview with a Nursing Assistant currently studying at Case Western University provides a more optimistic perspective. When asked if she foresaw a future in which nurses became obsolete due to technological advancements, she vehemently replied with a “no,” agreeing with my Section 3 conclusions. I briefly discussed the American Nurses Association's Code of Ethics Provision 4—which isn't legally binding but commonly used: nurses are ultimately accountable for their choices (American Nurses Association, “Code of Ethics” 31).

She believes that an interpretation of Provision 4 implies that nurses, while being able to use AI technologies, should retain the final decision, pointing to the historical usage of nurses of having decision-making power and ultimate judgment: “Different patients need different medication. For example, in breakthrough pain, patients need a different medication. And some medication needs multiple people [to] sign off. For opioids, I need another nurse to sign on it. I think it is completely impossible for AI to fully replace nurses because there will always be someone to sign off on it.” As a nursing assistant and prospective nurse, she would strongly prefer to “trust [her] own judgment than a robot,” and believed there would always be “skepticism around AI technologies.”

Although emphasizing that she was not an AI expert, he also doubted the capability of AI tools over nurses in certain scenarios. For example, when it came to pain recognition, she was doubtful about AI technologies such as computer vision (facial recognition). “AI can measure pain and discomfort, but say for example someone is going through chronic pain. AI would not recognize pain from the patient, because the patient will not be making a face of pain.”

She too pushed the concept that a robot worker would be “uncomfortable,” whereas a nurse may be more “adaptive,” “responsive,” and provide more psychological support, having certain “intuitive” and “emotional responsiveness” abilities that can receive emotional feedback from people in a manner that machines cannot. Further, machines may struggle with “fuzzy logic”, whereas nurses can better determine more broad and abstract factors. For example, nurses may be better at analyzing Social Determinants of Health (SDoH)—healthcare access and quality, education access and quality, social and community context, economic stability, and neighborhood and built environment—that current technologies would struggle to analyze.

She cites another example: “What if someone, maybe from a low-income environment, it is not safe for him to go home?” A nurse, she argues, would be able to recognize signs of discomfort from the patient that may be evidence of domestic violence. An AI might not be able

to provide the necessary help, as it wouldn't be trained to recognize these signs, or effectively capture the nuance of a situation.

Finally, she brings up certain concerns about efficiency, listing another example: “If someone has an injury to the bottom of the thigh, for example, you need multiple people to make sure that the patient is in the proper positioning, multiple types of adhesives.” As it is impossible to completely predict specific actions and scenarios, adaptiveness can be compounded with multiple workers, while having a singular machine may be less effective.

Her arguments against the obsolescence of nurses seem to be supported by data. The PEW research center finds that 60% of Americans would feel uncomfortable with their healthcare provider using AI technology (Tyson). Since one of the primary jobs of nurses is to make their patients more comfortable, this number may be more drastic for nursing AI.

Nevertheless, the people do rule in favor of complementarity in general (but not by a large margin). The PEW also finds that 38% would think that AI would lead to better health outcomes, 27% think there would be no difference, and 33% think it would become worse. In addition, the human element that nurses provide is also highly valued. The same study finds that 44% of people would not apply to a job that used AI to screen new hires, citing a lack of “human element” within their decision. In healthcare, surveyees cited damage to personal relationships between healthcare providers and customers as a primary concern (Faverio and Tyson).

Not only does she argue that nurses will not become obsolete, but she also disagrees with the idea that nursing demand may decrease, although she too foresees a high amount of complementarity, especially for smaller tasks, like simple Band-Aid application and even larger tasks, like creating treatment plans. From her point of view as an industry worker, she also contests the idea that the nursing occupation will undergo a drastic change, instead arguing that it would largely “stay the same.” This point of view, too, is reasonable. Productivity increases may not lead to layoffs, instead helping to fix the growing nurse staffing crisis (American Nurses Association, “Nurse Staffing Crisis”). “I’m not worried about nurse layoffs, there is always a shortage of nurses. If a nurse gets fired, it's not due to money issues.”

When asked if she envisioned a future where human interaction becomes devalued, she, although previously expressing some concern about “depersonalization” argue that human presence would always be “prevalent” in healthcare provision. She believes that between the chance to expand markets, or decreasing operating margins, companies will “definitely” choose to expand, and is “not all worried” about job prospects.

Indeed, while the analysis in Section 5 considers healthcare to be largely profit-driven, it also ignores healthcare incentives, especially ones targeting places that lack coverage. And even while considering healthcare to be largely profit-driven, increasing overall profits may be a more appealing choice than simply increasing profit margins. Furthermore, being early to expand into less-saturated markets, even at the cost of higher operating costs, can provide first-movers advantage, opportunities for diversification in products and customer bases, and provide benefits through the economics of scale. With the offer of increased productivity and cost advantage, AI technology may spur for-profit healthcare to make larger investments and expand.

She contests me on some points I previously identified as substitutable. Regarding the usage of AI in communication, she brings up a gadget used by nurses named Vocera. The tool is somewhat similar to a “radio phone.” Within inpatient situations, or other “case-by-case scenarios” that often become urgent, fast-paced communication and organization are required. These details have been included within Section 4, AI on Nursing. Overall, this interview demonstrates a more optimistic point of view. She believes that ultimately, nurses must hold the ultimate choice in care, and thus, cannot become obsolete. At the same time, she points to the limitations of AI technology, current and near-future, in rapid, quick-response situations, the belief in non substitutable humanity, and a constant demand for new nurses.

7 Conclusion

This study predicts the effects of current and near-future AI technologies on the nursing labor market. Within a precarious labor market, AI’s ability as a disruptor could bring large unemployment and job disruption for nurses, but is also capable of solving many problems, such as reducing burnout from large workloads. An analysis depicts nursing as a high-exposure, medium complementary occupation. At the same time, I argue that previously impervious tenets of human nurse healthcare, such as empathy, are more vulnerable than expected due to the ELIZA effect and affective computing. Despite rather high complementarity, in saturated markets with little opportunity for growth, higher productivity may cause job displacement. Unsaturated markets may lack the digital infrastructure to accommodate AI and may not see expansion due to high investment costs, but the predicted disruptive effects of AI may spur for-profits to reap the benefits of early-mover advantage and diversification. In addition, AI is predicted to have a transformative effect, causing task polarization.

This paper addresses the highly relevant issue of AI technologies, and their potential to replace occupations. However, instead of focusing on the labor market as a whole, I narrow the focus down to investigate the effects of AI on nursing. Occupations heavily implicated in AI substitution include software developers, data analysis, and even artists, but AI is anticipated to have a comprehensive effect. Thus, this case study may provide a distinctive analysis of an important yet relatively inconspicuous occupation.

Due to the highly speculative or sensationalist nature of AI technology, however, the analysis should be approached with caution. I accommodate for limitations of a potentially flawed framework that treats each task as equally labor intensive, but future studies can explore more nuance, such as the effects of humanitarian concerns, general health initiatives, or policies, and the effects of AI on specific types of nurses (as I use an umbrella term for nurses).

Works Cited

- Amaglobeli, David , et al. “Transforming Public Finance through GovTech.” IMF, IMF Publications, 7 Sept. 2023,
www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2023/09/06/Transforming-Public-Finance-Through-GovTech-535765. Accessed 24 July 2024.
- American Nurses Association. “Code of Ethics for Nurses.” Nursing World, ANA, Jan. 2015,
www.nursingworld.org/practice-policy/nursing-excellence/ethics/code-of-ethics-for-nurses/. Accessed 24 July 2024.
- . “Nurse Staffing Crisis.” Nursing World, ANA, 2019,
www.nursingworld.org/practice-policy/nurse-staffing/nurse-staffing-crisis/. Accessed 24 July 2024.
- Barr, Kyle. “Eating Disorder Helpline Takes down Chatbot after Its Advice Goes Horribly Wrong.” Gizmodo, 31 May 2023, gizmodo.com/ai-chatbot-eating-disorder-helpline-neda-1850490751. Accessed 13 Aug. 2024.
- Bresnick, Jennifer. “Artificial Intelligence in Healthcare Spending to Hit \$36B.” HealthITAnalytics, HealthITAnalytics, 28 Dec. 2018, healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b. Accessed 24 July 2024.
- British Council. “Should Robots Be Citizens? | British Council.” Www.britishcouncil.org, British Council Worldwide,
www.britishcouncil.org/anyone-anywhere/explore/digital-identities/robots-citizens. Accessed 24 July 2024.
- Brogan, Jen. “World Health Organization Launches AI-Powered Digital Health Assistant for Public Health - PMLiVE.” Pmlive.com, PMLive, 5 Apr. 2024,
pmlive.com/pharma_news/world-health-organization-launches-ai-powered-digital-health-assistant-for-public-health/. Accessed 25 July 2024.
- Brozen, Yale. Automation and Jobs. Selected Papers No. 18, 12 May 1965.
- Cazzaniga, Mauro, et al. “Gen-AI: Artificial Intelligence and the Future of Work.” IMF, IMF Publications, 14 Jan. 2024,
www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379?cid=bl-com-SDNEA2024001. Accessed 19 July 2024.
- Cognition Labs. “Introducing Devin, the First AI Software Engineer.” Www.youtube.com, Cognition, 12 Mar. 2024, www.youtube.com/watch?v=fjHtjT7GO1c. Accessed 19 July 2024.
- Cuadros, Diego F., et al. “Assessing Access to Digital Services in Health Care—Underserved Communities in the United States: A Cross-Sectional Study.” Mayo Clinic Proceedings: Digital Health, vol. 1, no. 3, 1 Sept. 2023, pp. 217–225,
www.sciencedirect.com/science/article/pii/S2949761223000305, <https://doi.org/10.1016/j.mcpdig.2023.04.004>. Accessed 25 July 2024.
- Fakhr-Movahedi, Ali, et al. “Exploring Nurse’s Communicative Role in Nurse-Patient Relations: A Qualitative Study.” Journal of Caring Sciences, vol. 5, no. 4, 1 Dec. 2016, pp. 267–276, www.ncbi.nlm.nih.gov/pmc/articles/PMC5187547/,
<https://doi.org/10.15171/jcs.2016.028>. Accessed 24 July 2024.
- Faverio, Michelle, and Alec Tyson. “What the Data Says about Americans’ Views of Artificial Intelligence.” Pew Research Center, 21 Nov. 2023,
www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/. Accessed 1 Sept. 2024
- Floud, Roderick, and D. N. McCloskey. The Economic History of Britain since 1700: Volume 1, 1700-1860. Cambridge [England] : Cambridge University Press, 18 Aug. 1994.

- Franklin University. “Bookkeeper vs. Accountant: Comparing Career Paths and Growth Potential | Franklin University.” www.franklin.edu, Accounting MVP Blog, 2021, www.franklin.edu/blog/accounting-mvp/bookkeeper-vs-accountant. Accessed 19 July 2024.
- Gao, Junfeng, et al. “Computer Vision in Healthcare Applications.” *Journal of Healthcare Engineering*, vol. 2018, 2018, pp. 1–4, <https://doi.org/10.1155/2018/5157020>.
- HMH. “HMH Acquires Award-Winning Software Company Writable, Launches Emerging Technology Incubator HMH Labs | HMH.” www.hmhco.com, Houghton Mifflin Harcourt, 28 Feb. 2024, www.hmhco.com/about-us/press-releases/hmh-acquires-award-winning-software-company-writable-launches-emerging-technology-incubator-hmh-labs. Accessed 25 July 2024.
- HRSA. “Nurse Workforce Projections, 2020-2035.” HRSA, Health Resources and Services Administration, Nov. 2022, bh.w.hrsa.gov/sites/default/files/bureau-health-workforce/Nursing-Workforce-Projections-Factsheet.pdf. Accessed 19 July 2024.
- . “Nurse Workforce Projections, 2021-2036.” HRSA, Health Resources and Services Administration, Mar. 2024, bh.w.hrsa.gov/sites/default/files/bureau-health-workforce/data-research/nursing-projections-factsheet.pdf. Accessed 11 Aug. 2024.
- International Labour Office. *ISCO-08 Vol. 1: International Standard Classification of Occupations*. International Labour Office, 2012. Accessed 1 Sept 2024.
- Lobdell, Nicole. “December 2023 Challenger Report: Job Cuts Fall to Second-Lowest Monthly Total of the Year; Job Cuts Steady as Hiring Plummet.” *Challenger, Gray & Christmas, Inc.*, 4 Jan. 2024, www.challengergray.com/blog/december-2023-challenger-report-job-cuts-fall-to-second-lowest-monthly-total-of-the-year-job-cuts-steady-as-hiring-plummet/. Accessed 20 July 2024.
- . “WSJ: Move to a New City for Work? No Thanks.” *Challenger, Gray & Christmas, Inc.*, 7 June 2021, www.challengergray.com/blog/wsj-move-to-a-new-city-for-work-no-thanks/. Accessed 25 July 2024.
- Nayak, Pandu. “Understanding Searches Better than Ever Before.” *Google*, 25 Oct. 2019, blog.google/products/search/search-language-understanding-bert/. Accessed 19 July 2024.
- OpenAI. “Introducing Sora — OpenAI’s Text-To-Video Model.” www.youtube.com, OpenAI, 17 Feb. 2024, www.youtube.com/watch?v=HK6y8DAPN_0. Accessed 19 July 2024.
- . “Say Hello to GPT-4o.” *YouTube*, OpenAI, 13 May 2024, www.youtube.com/watch?v=vgYi3Wr7v_g. Accessed 20 July 2024.
- Pailaha, Aprianto Daniel . “The Impact and Issues of Artificial Intelligence in Nursing Science and Healthcare Settings.” *SAGE Open Nursing*, vol. 9, 1 Jan. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10492460/, <https://doi.org/10.1177/23779608231196847>. Accessed 1 Sept. 2024
- Paris, Martine. “Snapchat Star Earned \$71,610 upon Launch of Sexy ChatGPT AI, Here’s How She Did It.” *Forbes*, Forbes, 12 May 2023, www.forbes.com/sites/martineparis/2023/05/11/carynai-virtual-date-earned-70000-with-sexy-chatgpt-ai-heres-how/. Accessed 25 July 2024.
- Petek, Gabriel. “The 2022-23 Budget: Health Care Access and Affordability.” *Legislative Analyst’s Office, LAO*, Feb. 2022, lao.ca.gov/reports/2022/4560/Health-Care-Access-Affordability-022322.pdf. Accessed 18 July 2024.

- Robert, Nancy. "How Artificial Intelligence Is Changing Nursing." *Nursing Management*, vol. 50, no. 9, Sept. 2019, pp. 30–39, www.ncbi.nlm.nih.gov/pmc/articles/PMC7597764/, <https://doi.org/10.1097/01.numa.0000578988.56622.21>. Accessed 24 July 2024.
- Rosseter, Robert. "Nursing Shortage Fact Sheet." *American Association of Colleges of Nursing, AACN*, May 2024, www.aacnursing.org/news-data/fact-sheets/nursing-shortage. Accessed 19 July 2024.
- Sull, Donald, and Charles Sull. "The Real Issues Driving the Nursing Crisis." *MIT Sloan Management Review*, 18 Oct. 2023, sloanreview.mit.edu/article/how-solve-nursing-crisis/. Accessed 1 Sept. 2024.
- Sun, Carolyn, and Kenrick Cato. "How Much Time Do Nurses Spend Using Electronic Devices at Work?" *Nursing Management*, vol. 51, no. 3, Mar. 2020, pp. 22–29, <https://doi.org/10.1097/01.NUMA.0000651184.19361.4e>. Accessed 1 Sept. 2024.
- The National Archives. "Why Did the Luddites Protest?" *The National Archives, The National Archives*, www.nationalarchives.gov.uk/education/resources/why-did-the-luddites-protest/. Accessed 25 July 2024.
- Toloka. "The History, Timeline, and Future of LLMs." *Toloka.ai, Toloka*, 26 June 2023, toloka.ai/blog/history-of-llms/. Accessed 19 July 2024.
- NU. "120 Healthcare and Nursing Statistics for 2024 | NU." *National University, NU*, 16 Mar. 2024, www.nu.edu/blog/healthcare-nursing-statistics/. Accessed 20 July 2024.
- Tyson, Alec, et al. "60% of Americans Would Be Uncomfortable with Provider Relying on AI in Their Own Health Care." *Pew Research Center Science & Society*, 22 Feb. 2023, www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/. Accessed 24 July 2024.
- Walker, Angelina. "Nursing News | Nurse.org." *Nurse.org, nurse.org*, 24 Apr. 2024, nurse.org/news/nurses-protest-ai. Accessed 25 July 2024.
- Wartzman, Rick. "The First Time America Freaked out over Automation." *POLITICO Magazine*, 30 May 2017, www.politico.com/magazine/story/2017/05/30/rick-wartzman-book-excerpt-automation-donald-trump-215207/. Accessed 18 July 2024.
- Webb, Michael. "The Impact of Artificial Intelligence on the Labor Market." *papers.ssrn.com*, 6 Nov. 2019, papers.ssrn.com/sol3/papers.cfm?abstract_id=3482150. Accessed 1 Sept. 2024.
- Weizenbaum, Joseph. "ELIZA---S Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM*, vol. 9, no. 1, 1 Jan. 1966, pp. 36–45, <https://doi.org/10.1145/365153.365168>. Accessed 1 Sept. 2024.
- Wootton, Charles, and Barbara Kemmerer. "1880-1930 Source: The Accounting Historians Journal." *The Accounting Historians Journal*, vol. 34, no. 1, June 2007, pp. 91–124, www.jstor.org/stable/40698352. Accessed 19 July 2024.
- Zarifhonarvar, Ali. "Economics of ChatGPT: A Labor Market View on the Occupational Impact of Artificial Intelligence." *Journal of Electronic Business & Digital Economics*, vol. 3, no. 2, 5 Dec. 2023, <https://doi.org/10.1108/jebde-10-2023-0021>. Accessed 1 Sept. 2024.
- Zilber, Ariel. "Tech Exec Predicts 'AI Girlfriends' Will Create \$1B Business." *New York Post, New York Post*, Accessed 15 Apr. 2024, nypost.com/2024/04/15/business/tech-exec-predicts-ai-girlfriends-will-create-1b-business/. Accessed 25 July 2024.

Implementation of Grover's Search Algorithm in Quality of Service Path Selection for the Border Gateway Protocol By Shashank Garag

Abstract

This quantitative research project aimed to improve the path selection of the Border Gateway Protocol (BGP) by incorporating network quality metrics. Due to the potential increase in computational intensity, it was tested whether implementing quantum algorithms could lower this factor while increasing network quality in path selection. The Quantum Minimum Search (QMS) algorithm, a form of Grover's Search Algorithm, was programmed in Python. Each part of the quantum circuit had a stored function, only being executed when necessary. This automated version of QMS that could be applied to any dataset was called Adaptive QMS (AQMS). The AQMS algorithm was used on a simulated BGP topology to determine the paths with the lowest delay, and highest bandwidth, with the AS-path prepending and ping tools used to alter path selection and measure throughput and latency. The average latency was reduced by 88.5% whereas the average throughput was increased by 1165.82%, for this specific topology. When comparing the time complexity of AQMS with classical linear search, there was a 54.70% reduction. A conducted t-test on both the throughput and latency results for BGP and AQMS BGP revealed that both p-values ($4.38 * 10^{-5}$ for latency and 0.00487 for throughput) were lower than the set significance level of 0.05, indicating that the null hypothesis was rejected, and AQMS was able to improve network quality while maintaining low computational intensity. This research provides a promising outlook for the application of quantum computers in the networking industry, as it increases the efficiency of data transmission, and provides a way to enhance the scalability of the internet and provide internet access to countries that cannot currently afford it.

1. Introduction

1.1 Border Gateway Protocol

The internet is divided into multiple sub-networks called autonomous systems (ASes), with connections between and within each AS for data packet transmission. The role of the Border Gateway Protocol (BGP) is to act as the primary facilitator of these ASes and determine the best path for a data packet to take to reach its destination. To do so, BGP has to factor in multiple path attributes and comply with any policies set by the network administrator of each AS. There are two specific types of BGP, internal BGP (iBGP), and external BGP (eBGP). Internal BGP focuses on routing a data package within the same network or AS, whereas eBGP focuses on routing a data package between different ASes. In this paper, the application will pertain solely to eBGP. Additionally, BGP is a path vector protocol, meaning it can track the entire path of a data packet to its destination through a series of AS numbers, allowing it to judge the shortest path to a packet's destination (Karteri 7). The list of path attributes used in BGP path selection is shown in Table 1.

Table 1: A list and order of BGP path attributes used during path selection. The BGP path selection algorithm checks to optimize each of the following path attributes, some of which are set by the network administrator before deciding on a path to route a packet.

Order	Path Attribute
1	Weight
2	Local Preference
3	Originate
4	AS Path Length
5	Origin Code
6	Multi-Exit Discriminator (MED)
7	eBGP vs. iBGP
8	Shortest IGP metric to Next Hop
9	BGP Multipath
10	Oldest Path
11	Lowest Router ID
12	Minimum Cluster List Length
13	Lowest Neighbor Address

BGP routers maintain connections by sending messages between each other, ensuring the session between them is still active. To share the path attribute information for the known paths in the router, an UPDATE message is sent to BGP neighbor routers in the form of routing tables. Routing tables store all known paths, their IP destination prefixes, and their corresponding path attributes such as weight and local preference. BGP routers compare the advertised routing tables to their current known paths to determine if the advertised paths are more optimal. An example of a BGP network, with connections, is shown in Figure 1.

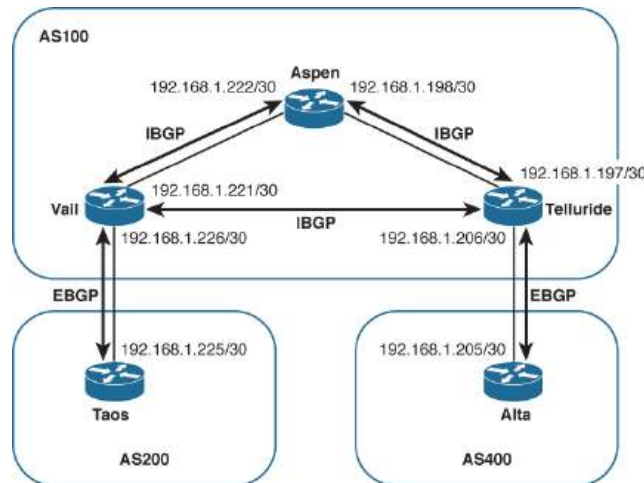


Fig. 1: A diagram of eBGP and iBGP connections across and within multiple ASes. The BGP path selection protocol determines the optimal path to send data packets from one router to another (Cisco Press).

However, even after considering a diverse set of path attributes, BGP still fails to incorporate Quality of Service (QoS) metrics into its path selection algorithm, prioritizing reachability across the internet over end-to-end network quality and performance, often leading

BGP to select paths that have high congestion or delay, drastically reducing the quality of network experienced by users on the internet. Some important QoS metrics that have a substantial effect on the experience of users are factors such as bandwidth, throughput, latency, and packet loss.

There have been efforts to allow BGP to accommodate QoS metrics in their path selection algorithm. A promising extension to the BGP path selection process that includes consideration of the path's QoS metrics is the Improved Stable Path Selection (ISPS) algorithm. ISPS builds off a previously created Stable Route Selection (SRS) algorithm that had failed to competently consider QoS metrics in the BGP path selection process. ISPS increases the efficiency of the BGP path selection process by incorporating the QoS metrics of delay and available bandwidth to find the most optimal path during the Adj-RIB-IN stage of the BGP path selection process. The results of the ISPS paper reveal that throughput increased by approximately 46.2% and almost no packet loss, indicating significant improvement in the SRS algorithm and the default BGP path selection process (Shukla and Kumar 16).

Although this algorithm seems promising in terms of improvements in network quality, it is still limited by the capabilities of classical computing. Routers are still tasked with sorting through thousands of available paths to determine the most optimal path. This procedure can often be taxing on routers and computers, especially if they are instructed to find the best QoS metrics in addition to the default BGP path selection algorithm. For this reason, the ISPS algorithm is limited to only comparing two paths at a time. However, quantum computing has shown success in finding specific values in large databases at a lower complexity level than classical computing. This is promising for finding the most optimal path based on QoS metrics, in large routing tables. Therefore, the application of quantum computing algorithms could potentially increase the efficiency of BGP path selection, and decrease the computational time necessary to find the best path.

In this research paper, quantum computing techniques will be implemented in the BGP path selection process, specifically the routing decisions made in eBGP, to identify the paths with the most optimal QoS metrics of available bandwidth and delay. The main quantum algorithm implemented in this paper is Grover's Search Algorithm, which is designed to find a specific value in an unstructured database. However, because the value of the most optimal metric is unknown, a specific form of Grover's Search Algorithm will be used, known as the Quantum Minimum Search (QMS) algorithm. An improved version of the QMS algorithm, for a networking application, will be utilized in the Cisco Packet Tracer network simulator to extract results regarding the effects of the algorithm on the end-to-end network quality of eBGP. This paper hypothesizes that the implementation of the QMS algorithm in the routing process of eBGP will reduce the convergence time of eBGP routers, and improve end-to-end network quality in the form of increased throughput and reduced packet loss.

1.2 Purpose

The purpose of this quantitative, experimental project is to test the hypothesis that the implementation of the QMS algorithm in eBGP routing processes enables routers to consider additional QoS metrics in their path selection decisions, improving the quality of network experienced by users while maintaining a low convergence time during path selection. This experiment will analyze the differences in computational complexity of classical and quantum algorithms when they are used to find the most optimal path in a routing table, and how each defines the capability of a router to improve the end user network quality through the consideration of QoS metrics in addition to the path attributes currently present in the BGP path selection process. The potential improvements in network quality could speed up daily interactions between different networks and internet service providers, ultimately making everyday tasks for people in society much faster. Issues such as the crashing of web pages due to traffic could be better avoided.

1.3 Structure

The rest of the paper is structured as follows: Section 2 covers the fundamentals of quantum computing, Grover's Search Algorithm, and Dürr and Høyer's Algorithm. Section 3 covers the Quantum Minimum Search algorithm, and how it implements Grover's Search Algorithm to accomplish its goal. Section 4, covers the methodology of the project, an overview of the Cisco Packet Tracer simulator, and the implementation of the QMS algorithm in the simulator. Section 5 provides a summary of the results extracted from the simulator and an analysis of their improvements from the default BGP path selection algorithm. Section 6 ends the paper with a summary of the experiment and its potential implications for the future of quantum computing applications in networking.

2. Quantum Computing

2.1 Essentials

Unlike classical computing, which utilizes bits to represent and store information, quantum computing uses qubits, which follow the rules of quantum mechanics. In classical computers, bits store information in the form of 1's or 0's, with a 1 indicating true, and a 0 indicating false. Qubits, however, can represent information in the states of 1 and 0, but also as a combination of both of these states. This allows quantum computers to explore various possibilities at once, reducing the computational time needed to solve certain problems, in comparison to classical computers. Another contrast to classical computers is the form of circuits that are implemented in quantum computers. Quantum computers use quantum circuits with quantum gates to perform necessary computations and structure algorithms, whereas, in classical computing, logic gates are used. Quantum gates allow quantum computers to take advantage of techniques based on quantum mechanics, such as superposition and entanglement.

Superposition is the concept of a qubit being imposed into a state that is a combination of the orthonormal basis states of 0 and 1. The state of a qubit can be represented as a vector, and they are written using the bra-ket notation: $|0\rangle$, $|1\rangle$. When a quantum gate is applied to a qubit in either basis state, it will be superimposed into a combination of both basis states. A quantum gate is simply a matrix, and applying it to a qubit means multiplying the matrix by the state vector of the qubit. When superimposed, the qubit has a certain probability of being measured as the $|1\rangle$ or $|0\rangle$ state. The specific probability of its measurement varies depending on the quantum gates used to superimpose the qubit. To give an example, the Hadamard gate, a common quantum gate that is used in many quantum circuits, including Grover's Search Algorithm, is shown in equation 1 below:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (1)$$

When the Hadamard gate is applied to the basis state of $|0\rangle$, it can be represented as the following matrix, which is also called the $|-\rangle$ state, another form of a basis state, shown in equation 2:

$$|q\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (2)$$

The other main advantage of quantum computing is entanglement. Entanglement is the concept of two qubits being intertwined through their measurement results. Regardless of the distance between them, measuring one of the entangled qubits would automatically provide the information of the other qubit's measurement result, without the need to measure the other qubit itself. An example of a quantum gate that can be used to establish entanglement between two qubits is the CNOT-gate. The CNOT-gate (Controlled NOT gate) is a conditional quantum gate, which applies the X-gate on a target qubit based on the state of the control qubit. If the control qubit is in the basis state 1, the CNOT will apply the X-gate on the target qubit. Both the CNOT-gate and the X-gate are depicted below in equations 3 and 4:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (3)$$

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (4)$$

Consider a quantum circuit with two qubits, with the first qubit being in the state $|+\rangle$ and the second in state $|0\rangle$. If a CNOT-gate was applied to these qubits, with the first qubit being the control and the second qubit being the target, it would change the original statevector (a vector representation of the superposition and entanglement of the two qubits shown in equation 5):

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (5)$$

into the following statevector in equation 6:

$$\left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \quad 0 \quad 0 \right] \quad (6)$$

with the second and last amplitudes being swapped.

Through the applications of quantum gates to perform superposition and entanglement, various quantum algorithms can be created (Williams 51-111).

2.2 Grover's Search Algorithm

The Grover's Search Algorithm is a specific quantum circuit that allows for a faster search of an unstructured database. Classical computing search algorithm would take $N/2$ tries to search the entire database for the desired value at best, however, using Grover's Search Algorithm, only \sqrt{N} tries are needed to find the desired value, resulting in a quadratic speed up. Grover's Search Algorithm works through three steps, the state preparation, the oracle, and the diffusion operator.

The first part is the state preparation, which superimposes all the qubits being used to represent the database, such that the probability of any one of the superimposed states being measured is equal. To create the state preparation operator, the Hadamard gate is applied to all qubits in the quantum circuit, equalizing the amplitudes of all the states. In a graphical representation, the desired state can be represented by a vector $|w\rangle$ and the superposition can be represented as such: $|s\rangle = H^{\otimes n} |0\rangle^n$.

The second part is Grover's Search Oracle, which is the most important part of Grover's Search Algorithm. The oracle is used to "mark" or identify the desired state the algorithm is looking for in the entire database. The oracle marks this desired state by applying a negative phase on the state's amplitude. Referring back to the graphical interpretation, the oracle would reflect the vector $|s\rangle$ across the vector $|s'\rangle$, which is simply a vector perpendicular to the desired vector $|w\rangle$. This makes it easier for the algorithm to differentiate the desired state from the other values in the database, which leads us to the next step of Grover's Search Algorithm, the diffusion operator.

The diffusion operator once again reflects the amplitude of the desired state, but it also magnifies the amplitude of the desired state and minimizes the amplitudes of the other states in the database. This process creates a drastic difference between the amplitudes of the desired state and the other states in the database, in turn increasing the probability of the desired state being returned after the superposition collapses from the measurement operator. In terms of the graphical interpretation, it would perform the following operation on the vector $|s\rangle$ (which has already been reflected once around $|s'\rangle$): $|s\rangle \langle s| - \mathbb{1}$ (Giri and Korepin 9–11).

The combination of the oracle and the diffusion operator can be defined as a single rotation. The algorithm is repeated for a certain number of rotations until the probability of the measurement collapsing into the desired state is almost guaranteed. Graphically this would mean that the vector $|s\rangle$ has reached the desired vector $|w\rangle$. Due to the much smaller amount of

iterations - or rotations - needed to return the desired value in the database, the computational time is reduced significantly, especially for large databases.

2.3 Grover Adaptive Search

Grover's Search Algorithm can be modified to perform the functionalities of an adaptive search algorithm, called Grover Adaptive Search (GAS). The main goal of GAS is to optimize an objective function, meaning it has to find the value in a database that provides the best output of a certain function, such as the minimum or maximum. Grover Adaptive Search alters Grover's Search Algorithm by changing the structure of its oracle to accommodate additional features. The main change to Grover's Search Algorithm is that GAS can keep track of the currently known best value in the database for every rotation. GAS uses the currently known value to find better solutions to the objective function at every rotation, which is useful when it is unknown to the user what the best solution in a database is. After a certain number of iterations, GAS is eventually able to find the dominant solution in a data set.

Another parameter that is necessary for the execution of GAS is the rotation count for every iteration of the Grover Search framework. Although it is possible to give any number for the rotation count, an optimal rotation count exists, and the implementation of the optimal rotation count can greatly increase the efficiency of GAS. To provide the optimal rotation count, another form of GAS exists, called Dürr and Høyer's Algorithm.

Dürr and Høyer's Algorithm adds a procedure to find the optimal rotation count for every iteration of GAS. First, an extra parameter, λ , is set to a certain value, preferably $8/7$. Then, the algorithm selects a random rotation count from a set of numbers from 0 to $m - 1$, with $m = 1$ initially. If a dominating solution is found, $m = 1$, otherwise m is set to the minimum of either λm or \sqrt{N} , as \sqrt{N} is the number of tries necessary to find the desired value using Grover's Search Algorithm (Baritompa et al. 6–9). The full algorithm for Dürr and Høyer's is shown below in algorithm 1:

Algorithm 1 Dürr and Høyer's Algorithm

- 1: Randomly choose x from the decision space.
 - 2: Set $x_1 \leftarrow x$
 - 3: Set $y_1 \leftarrow f(x_1)$
 - 4: Set $m \leftarrow 1$
 - 5: Choose a parameter λ (Preferably $\frac{8}{7}$)
 - 6: For $k = 1, 2, \dots$ until termination condition is met, do:
 - a. Choose a random rotation count r_k uniformly from $[0, \dots, [m - 1]]$.
 - b. Apply Grover's Search Algorithm with r_k iterations on $f(x)$ with the threshold y_k , and denote outputs as x and y .
 - c. If $y < y_k$, set $x_{k+1} \leftarrow x$, $y_{k+1} \leftarrow y$ and $m \leftarrow 1$; otherwise, set $x_{k+1} \leftarrow x_k$, $y_{k+1} \leftarrow y_k$ and $m \leftarrow \min\{\lambda m, \sqrt{N}\}$.
-

3. Quantum Minimum Search Algorithm

The QMS algorithm is an implementation of Dürr and Høyer's algorithm, as it implements Grover's Search algorithm as a subroutine, iteratively updating the desired value until it identifies the minimum value in the database. However, the unique aspect of the QMS algorithm is that, after applying superposition on all the indexes of the database (which are converted to binary), it uses the concept of QRAM (Quantum RAM) to store the values in a database in binary form. QRAM enables the algorithm to correspond the binary indexes of each value in the database with the binary values themselves. This correspondence is created through a series of multi-controlled X-gates.

After encoding all the values in the QRAM, an oracle is applied to the binary values. The oracle of the QMS algorithm allows it to follow the structure of Dürr and Høyer's algorithm, iteratively updating the desired or best value in the database. The oracle itself is a multi-controlled X-gate, however, the state of its control qubits changes depending on what the current best value is. During its first iteration, the oracle has one control qubit in the $|0\rangle$ state, which means it searches for all values in the database with the first binary digit of 0, as they are the smallest values. After this iteration, a random binary value with the first digit of 0 is picked. For the second iteration, another control qubit is added in the $|0\rangle$ state, and once again a binary value with the first digit of 0 is picked. However, this time, the second digit could either be a 1 or a 0. If it is a 0, it means that there exists a value in the database with a second digit of 0, however, if it is a 1 it means that there doesn't exist a value in the database with a second digit of 0. This is fine, as it indicates that the smallest value in the database has a second digit of 1. This iterative process continues until all the control qubits have been filled for the oracle, with their respective states. At the final iteration, the oracle can mark a single value, which is the smallest value in the database (Albino et al. 2–4).

As for all Grover Search algorithm circuits, every iteration ends with a diffusion operator, and a measurement of the qubits that encode the indices of the values in the database, leading the circuit to output the index of the value the oracle marked. There is also room for error in this

circuit however, because the selection of a value from the database by the oracle is probabilistic, and there isn't always a guarantee that the value marked by the oracle will be selected, even if it is the majority of the time. Therefore, instead of relying on a single value from the quantum circuit to reveal the path with the most optimal QoS metrics, every output of each iteration is stored in a list, and the classical minimum function in Python is used to find the most optimal path. This is not very computationally intensive, as, at the end of all the iterations, the list is small enough such that it requires barely any computational resources for a classical computer to find the minimum value in that list. The algorithm for QMS is shown below in algorithm 2:

Algorithm 2 Quantum Minimum Search Algorithm

- 1: Randomly choose y from the decision space, whose value needs m bits to be stored.
- 2: Initialize the quantum circuit in the state $|\psi_0\rangle = \frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x\rangle |0\rangle^{\otimes m} |-\rangle$
- 3: Store values in QRAM to get the state $|\psi_1\rangle = \frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x\rangle |y_x\rangle |-\rangle$
- 4: Apply Oracle P to mark all values with the least significant bit of 0, which are the first set of the smallest values in the database.
- 5: Apply a diffuser operator W on the marked states
- 6: Apply measurement to obtain a value y_{i+1} which is less than y_i
- 7: If all qubits have been analyzed end algorithm.
- 8: Else continue.
- 9: return y_i

4. Methodology

4.1 Tools

The tools used in this project to re-create the QMS algorithm were the Qiskit Software Development Kit, the Jupyter Notebook IDE, and the PyCharm IDE. The Qiskit SDK allowed for the implementation of quantum programming techniques as well as access to quantum simulators to run the program. The PyCharm and Jupyter Notebook IDE's were used to write the program in the Python programming language, with the Jupyter Notebook mainly used for better visualization of the circuit, and the probability histograms. The tool used in this project to set up the simulation of the Border Gateway Protocol was the Cisco Packet Tracer software, which provides the basic simulation functionality for networking with the BGP protocol.

The Cisco Packet Tracer simulator is software that allows one to create networks of routers and various internet devices. Connections can be made between these routers using various routing protocols, and data can be sent through these networks. Multiple different topologies can be created, and the resulting ping and latency of data transmission can also be measured. However, there are certain limitations to the Cisco Packet Tracer, as it doesn't allow for full functionality and manipulation of certain network protocols such as BGP. Due to the

limitations of the computer hardware used in this experiment, the Cisco Packet Tracer simulation was the only viable option for testing the quantum algorithm for BGP path selection.

4.2 Programming QMS Algorithm

To create a QMS Algorithm that can adapt to databases of different sizes with values of different magnitudes, an Object Oriented approach was taken. The entire program was written as a class, with each of the different parts of the circuit as different functions. Each function returned its respective part of the circuit. Finally, there was one solve function that composed together all the different parts (the QRAM, oracle, diffuser, etc.) and created the full quantum circuit for the QMS Algorithm. It also implemented the iterative process of measuring the circuit and updating the control qubits for the oracle, which was done by having a running list of the qubits to which the control qubits would be applied and another list that specified whether the state of the control qubit was $|0\rangle$ or $|1\rangle$. After the output of every iteration was stored in a list, the Python `min()` function revealed the smallest value in the database. The use of the `min()` function, which is a classical minimization function, is to overcome the probabilistic nature of Grover Search. Because there is a small chance that the wrong minimum value may be selected, even if the optimal value was selected in a previous iteration, the `min()` function allows for the correct answer to be selected at the end. Additionally, this operation does not add extensive time complexity to the quantum circuit, as the list of values outputted by the quantum circuit is much smaller than the original database, to an extent where the time complexity added is insignificant.

Additionally, a save-and-execute methodology was used to accommodate topology changes in the BGP simulator. As explained before, the main parts of the QMS circuit were created and returned by their respective functions. A separate function was implemented to store all the returned circuit objects as files so that they can be easily loaded and executed in the main solve function. This allows for a lower time complexity and fewer computational resources to be used when the BGP topology remains constant. It is also important for network administrators to avoid constant changes to the topology, as this could lead to excessive recompilation of the QMS circuit, and too much computational overhead. The algorithm for the adaptive QMS or AQMS algorithm is shown below in algorithm 3:

Algorithm 3 Adaptive Quantum Minimum Search Algorithm

- 1: Check if there is a topology change, and if so recompile the QRAM with the updated information:
 - a. Re-initialize the state preparation operator
 - b. Recompile the U-gate by looping through all the binary indexes and values and creating a series of multi controlled not gates.
 - c. Re-initialize the diffuser operator
 - d. Save all of these quantum components, as well as the variables used to track the iterative adaptive search and update the oracle.
 - 2: Else, continue as follows:
 - a. Randomly choose y from the decision space, whose value needs m bits to be stored.
 - b. Re-load all saved objects to be used for the following process.
 - c. Initialize the quantum circuit in the state $|\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle |0\rangle^{\otimes m} |-\rangle$
 - d. Store values in QRAM to get the state $|\psi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle |y_x\rangle |-\rangle$
 - e. Apply Oracle P to mark all values with the least significant bit of 0, which are the first set of the smallest values in the database.
 - f. Apply a diffuser operator W on the marked states
 - g. Apply measurement to obtain a value y_{i+1} which is less than y_i
 - h. If all qubits have been analyzed end algorithm.
 - i. Else continue.
 - 3: Return y_i
-

4.3 Testing Procedure and Simulation Set-Up

To verify the accuracy of the hypothesis, it was necessary to measure the time complexity of the quantum algorithm, implement the corresponding changes in the BGP path selection, and measure the resultant throughput and latency. The time complexity of the quantum algorithm was measured in two ways, first calculating the theoretical time complexity, which is represented in the big O notation, and second, testing the circuit on a quantum computer and measuring its execution time. In the BGP simulation, based on what the most optimal path was (the path with the lowest delay and highest available bandwidth), the technique of AS-Path prepending was used to modify the BGP routing tables and influence the BGP path selection to select the desired path. AS-Path prepending influences BGP path selection by adding multiple AS numbers to undesired paths, making it seem longer, and because BGP path selection prefers shorter paths, the paths altered by AS-Path prepending are not chosen (Chang and Lo 1). The throughput and latency results were compared between the normal BGP path selection, and the path selection influenced by QMS to verify improvement.

The topology was set up in the Cisco Packet Tracer with a total of five routers, and connections between each of them. They were all configured to run the BGP routing protocol and were paired with each other using the network and neighbor commands in the router's CLI. Then, the individual path's delays and bandwidths were set using their respective commands in each router's CLI. The labeled topology can be seen in Figure 2 below:

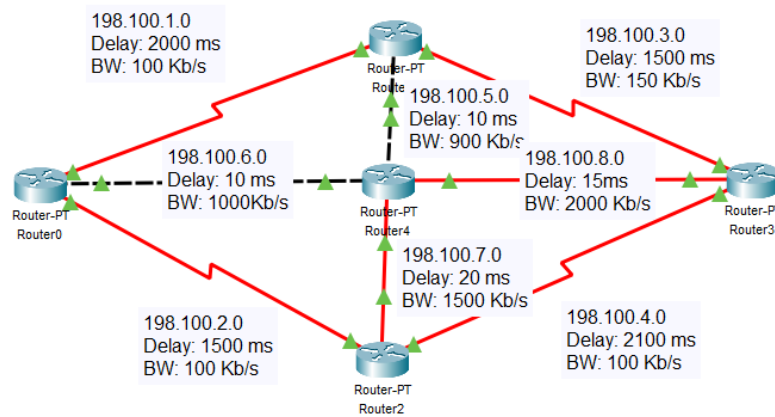


Fig. 2: The BGP topology used for testing. There were eight connections made between 5 different routers, each router representing an autonomous system. Each path had varying delay and bandwidth values, intended to change the resulting network quality based on which path the data packet was transmitted.

5. Results and Discussion

5.1 Average Latency

The ping command was utilized to measure the latency of a packet being transmitted to its destination and then returned to its sender. For the specific topology in this experiment, Router 0 sent Router 3 a packet transmission with varying packet sizes. For the default BGP path selection, the path across connections 198.100.1.0 and 198.100.3.0 was used to transmit the data packet. However, as seen in Figure 2, this is not the most optimal path as the delay is not the lowest, and the available bandwidth is not the highest. Thus, as seen in Figure 3, this led to latencies that were sub-optimal, and high, which would result in a lower end-user network quality, and longer wait times to execute processes on the internet. However, as determined by the AQMS algorithm, and by visually looking at the BGP topology, the path formed from connections 198.100.6.0 and 198.100.8.0 is the most optimal. It contains the lowest delay and the largest available bandwidth. By using the output of the AQMS algorithm, an automated Python program with a library such as Netmiko, the AS-Path prepending method can be implemented to

modify BGP to favor this path. When the ping was sent through this path, the average latency for all variations of packet sizes was much less than that when the original BGP path selection was used. There was a mean decrease of 88.5 percent in overall latency when the AQMS algorithm was used to determine the optimal path to send the data packet through, ultimately proving the original hypothesis, as the user can experience lower latency over an eBGP network, increasing their network quality.

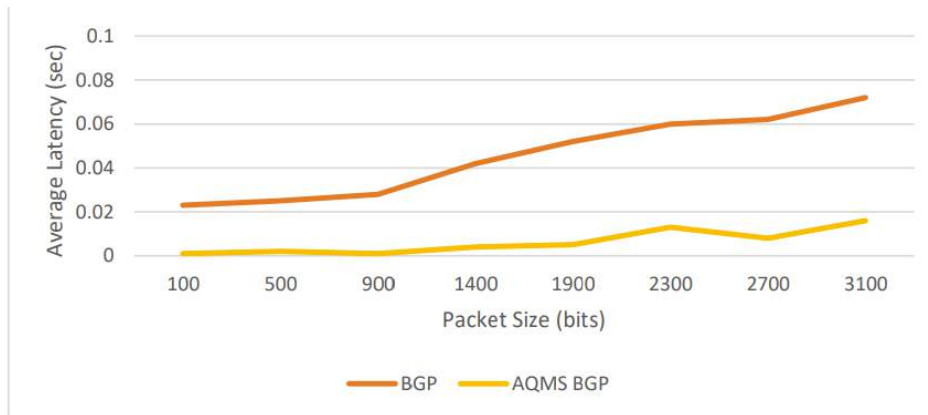


Fig. 3: A comparison of Average Latency for BGP with and without AQMS. The red line represents the default BGP path selection, which resulted in consistently higher latencies of packet transmission. The yellow line represents the path selection done by AQMS BGP, which had consistently lower latencies of packet transmission.

5.2 Average Throughput

The average throughput was determined through the previous extraction of the average latency through the BGP simulation. Since throughput is defined to be the amount of information sent per unit of time, by dividing the packet size by the average latency, the throughput was determined. This would be an accurate calculation, as the simulation consistently revealed that all packets were delivered without any packet loss for both the AQMS and non-AQMS implementation of the BGP path selection. When this calculation was made for all variations of packet sizes, it could be seen that compared to regular BGP routing, the implementation of AQMS to select the most optimal path (the path that was described in the previous subsection) led to greater overall throughput, as shown in Figure 4. In fact, through the calculation of the average, it was determined that throughput increased by 1165.82 percent. Although there was some variation in throughput, there was an overall downward trend as packet size increased. This thoroughly proves the hypothesis that the implementation of AQMS for path selection in eBGP networks leads to better network quality. A user on the internet would be able to send and receive more data in a shorter time frame, making their experience and network quality much greater.

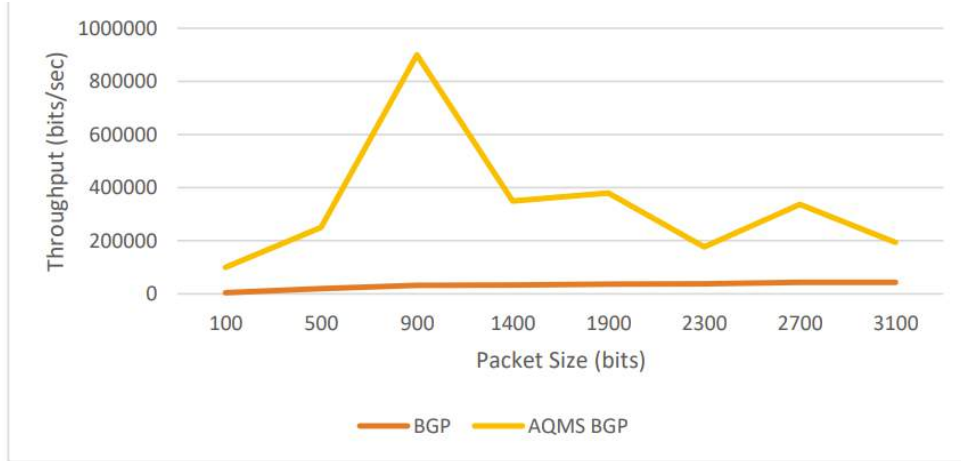


Fig. 4: A comparison of Average Throughput for BGP with and without AQMS. The red line represents path selection by default BGP, which has consistently lower throughput. The yellow line represents path selection done by AQMS BGP, which had consistently higher throughput.

5.3 Time Complexity

The time complexity of the improved QMS algorithm was measured using the big O notation, which is commonly used to measure the algorithmic complexity of programs. As explained previously in the background info, Grover's Search Algorithm is known to have a time complexity of $O(\sqrt{N})$. Specifically, for the implementation of a QRAM circuit to execute Grover Adaptive Search, the range of complexity was stated to be from $O(6\sqrt{N})$ to $O(14\sqrt{N})$. For a classical linear search, the time complexity would be $O(N)$, meaning it would have to go through all the values in the database before reaching the desired value, in a worst-case scenario. When quantifying the improvement that AQMS made over a classical algorithm in terms of time complexity, there was a 54.70% decrease in time complexity with AQMS. However, this time complexity varies due to the state of the networking topology. If the topology remains unchanged, the time complexity remains about the same for the execution of the program. On the other hand, if there is a sudden change in topology, the time complexity will change to be about $O(\sqrt{N} + N + N^3)$, which is a drastic increase in the execution time of the quantum algorithm. This is because the recompilation of the quantum gates and the QRAM involves an extensive set of loops in each part of the quantum circuit, such as the diffuser and the U gate. The U gate especially has a triple nested for loop, which leads to the N cubed term in the time complexity. In terms of a networking application, this would advise network operators managing a BGP network to avoid frequent changes in the topology, such as the link configuration and the bandwidth settings on each connection. However, the object-oriented design of the quantum circuit recompilation provides certain advantages in a networking application such as a BGP path selection algorithm. When storing the outputs of the defined functions, and loading them during the quantum circuit execution, minimizes the amount of times it is necessary to recompile each part of the quantum circuit. It streamlines the program to focus on the continuous execution of path selection, with rare instances where an entire update to the circuit is necessary. Thus, it can

be concluded that the improved QMS algorithm still provides a quadratic speedup over classical search algorithms.

5.4 Hypothesis Testing

To confirm the validity of the experiment's results, further statistical analysis was conducted to verify the null hypothesis was rejected. In this case, the null hypothesis was that using AQMS in BGP path selection would not affect the improvement of network quality. A t-test was conducted, with the significance level, α , being set to 0.050. When a one-tailed distribution t-test was conducted for both the output comparisons of throughput and latency, p values of 0.49% and $4.38 * 10^{-3}\%$ were obtained respectively. As both of these p-values were lower than the significance level, the null hypothesis was rejected, and the original hypothesis that AQMS provides an increase in network quality along with a lower time complexity was proved. The time complexity values for all trials are shown in Table 2 below.

Table 2: Comparison of time complexities for AQMS and classical linear search algorithm

N	Classical	AQMS	AQMS with Topology Change
100	100	100.00	1000110.00
200	200	141.42	8000214.14
300	300	173.21	27000317.32
400	400	200.00	64000420.00
500	500	223.61	125000522.36
600	600	244.95	216000624.49
700	700	264.58	343000726.46
800	800	282.84	512000828.28

6. Conclusion

This experiment ultimately proved the original hypothesis that implementing AQMS in BGP path selection not only allows for QoS metric accommodation, and overall better network quality, but also maintains a low time complexity, preventing excessive computational stress on the routers. The results revealed that when the output of AQMS was used to alter path selection through the process of AS-path prepending, both the latency of packet transmission was decreased and the overall throughput was increased. In terms of the bigger implications of these results, users and companies would be able to experience faster data transmission across eBGP networks, allowing for information to be acquired quickly, and allowing for more efficient execution of daily tasks. Additionally, an increase in network quality makes it easier to increase the Internet's scalability and makes it more accessible to people all over the world. For countries

that have difficulty providing internet access, the implementation of an algorithm that provides an increase in network quality could make it more affordable for these countries to give their citizens this essential resource. However, quantum computers are still developing, and can only handle processing a certain number of qubits, which may not be sufficient to address a large network. As research progresses in the field of quantum computing, and quantum computers can process more and more qubits, this experiment provides another potential application for the advantages of quantum algorithms.

Some aspects of this research are still yet to be addressed and are spots for future improvement. For example, the fact that a network topology change leads to such a large time complexity can be an issue for BGP convergence, which is already slow as it is. If this issue can be resolved, either through methods to fix the large amount of pre-processing or the eventual advancements in quantum computers, it would allow for a consistently low time complexity for AQMS in BGP path selection. A potential solution to avoid the high time complexities when there are topology changes is to manipulate the quantum computing to identify where the topology change occurred (which specific path or node) and update that part of the quantum circuit in real-time. This would prevent the quantum circuit from having to be recompiled fully, and this can be done through machine learning techniques to identify changes in the network topology. Additionally, although AS-Path prepending is a good way to alter BGP path selection, it would be better if AQMS could be more directly integrated into the BGP path selection process. This would reduce the amount of stages the program would have to go through until it executes its final decision, reducing the overall runtime of the implementation.

Works Cited

- Albino, Anton S., et al. "Quantum Algorithm for Finding Minimum Values in a Quantum Random Access Memory." arXiv.org, 12 Jan. 2023, arxiv.org/abs/2301.05122.
- Baritompa, W. P., et al. "Grover's Quantum Algorithm Applied to Global Optimization." SIAM Journal on Optimization, vol. 15, no. 4, Jan. 2005, pp. 1170–84. <https://doi.org/10.1137/040605072>.
- Chang, Rocky K. C., and Michael Lo. "Inbound Traffic Engineering for Multihomed ASs Using AS Path Prepending." IEEE Journals & Magazine | IEEE Xplore, 1 Apr. 2005, ieeexplore.ieee.org/abstract/document/1407694.
- Configuring and Troubleshooting BGP Peering > Introduction to BGP | Cisco Press. www.ciscopress.com/articles/article.asp?p=2738462&seqNum=3.
- Giri, Pulak Ranjan, and Vladimir E. Korepin. "A Review on Quantum Search Algorithms." Quantum Information Processing, vol. 16, no. 12, Nov. 2017, <https://doi.org/10.1007/s11128-017-1768-7>.
- Karteri, Ledina Hoxha. "Bgp: A Protocol for the 'Inter Domain' Routing." Questa Soft, 2017, www.cceol.com/search/article-detail?id=1238964.
- Shukla, Shipra, and Mahesh Kumar. "An Improved Quality Path Selection Approach for Border Gateway Protocol." International Journal of Intelligent Engineering and Systems, vol. 10, no. 5, Oct. 2017, <https://doi.org/10.22266/ijies2017.1031.02>.
- Williams, Colin P. "Explorations in Quantum Computing." Google Books, 2011, books.google.com/books?hl=en&lr=&id=QE8S--WjIFwC&oi=fnd&pg=PR8&dq=C.P.+Williams.+Explorations+in+Quantum+Computing.+Springer-Verlag+Lon-+don+Limited,+2011.&ots=BMEMUR_xXI&sig=vQNPDJWtk46mRX1jk4R6gYq73sQ.

The Relationship Between Video Game Use and Attention-Hyperactivity Disorder (ADHD) Among Children and Adolescents
By Oruebubechukwu Asogwa

Abstract

Video games have become popular entertainment among children and adolescents worldwide. Excessive video game play, with its reinforcement of sensation-seeking and instant gratification, may be appealing to individuals with attention-deficit/hyperactivity disorder (ADHD) and can worsen ADHD symptoms. Children with ADHD are predominantly inattentive, impulsive, and lack self-control, leaving them more susceptible to excessive video game use when compared to children without ADHD. This systematic review aimed to examine the relationship between video game use and ADHD among children and adolescents.

PubMed, Embase, PsycINFO, and CINAHL were searched using the following terms: *children, adolescents, video games, and ADHD*. Based on the established inclusion criteria, 18 articles were selected for extraction. Overall, the results showed that children diagnosed with ADHD are more susceptible to excessive video game use. Video games can exacerbate ADHD symptoms, but two of the reviewed studies found no relationship between video game use and ADHD.

This review aimed to evaluate the relationship between ADHD and video game use among children and adolescents. Findings revealed that children with ADHD are at a higher risk of becoming addicted to video games, and video game use exacerbated ADHD symptoms. In addition, comprehensive and uniform scales for ADHD and video game use need to be developed to reduce the variability in the findings of the studies. More research needs to be conducted on the relationship, primarily among underrepresented regions such as the Middle East, South America, and Africa.

Keywords: ADHD, video game use, children, adolescents, systematic review

Introduction

Video games are popular entertainment among children and adolescents (Muzwagi et al., 2021). According to Kids and Gaming (2011), over 90% of children and adolescents play video games and spend significant time playing them (NPD Group, 2011; Gentile et al. 2017). This increase in video game time and use has led to public concerns about potentially damaging effects such as obesity, anxiety, depression, aggression, and poor school performance, coupled together with the possibility that video game play may become addicting to children and adolescents (Brown and Witherspoon 2002; Alrahili et al., 2023). Research has revealed that excessive video game play, reinforcing sensation-seeking and instant gratification, may be appealing to individuals with attention-deficit/hyperactivity disorder (ADHD) and may worsen ADHD symptoms (Masi et al., 2021). Due to the rise in teen consumption of online video games

and their potential risk factors, an inquiry into the directionality and relationship between ADHD and video game use among children and adolescents is warranted.

According to the World Health Organization (WHO), video games, whether played online or offline, have been linked to mental health problems in children and adolescents (World Health Organization, 2019). The 11th Revision of the International Classification of Diseases (ICD-11) defines *gaming disorder* "as a pattern of gaming behavior characterized by impaired control over gaming, increasing priority given to gaming over other activities to the extent that gaming takes precedence over other interests and daily activities, and continuation or escalation of gaming despite the occurrence of negative consequences" (World Health Organization, 2023)

ADHD is a behavioral condition that develops before the age of 12 and involves the inability to sustain attention and focus on tasks; that is, people with ADHD struggle with being organized, completing assignments, staying focused, and thinking before acting (National Institute of Mental Health, 2023). Symptoms of ADHD include restlessness, inattentiveness, poor impulse control, defiance, social ineptness, or aggression (Hedges et al., 2013).

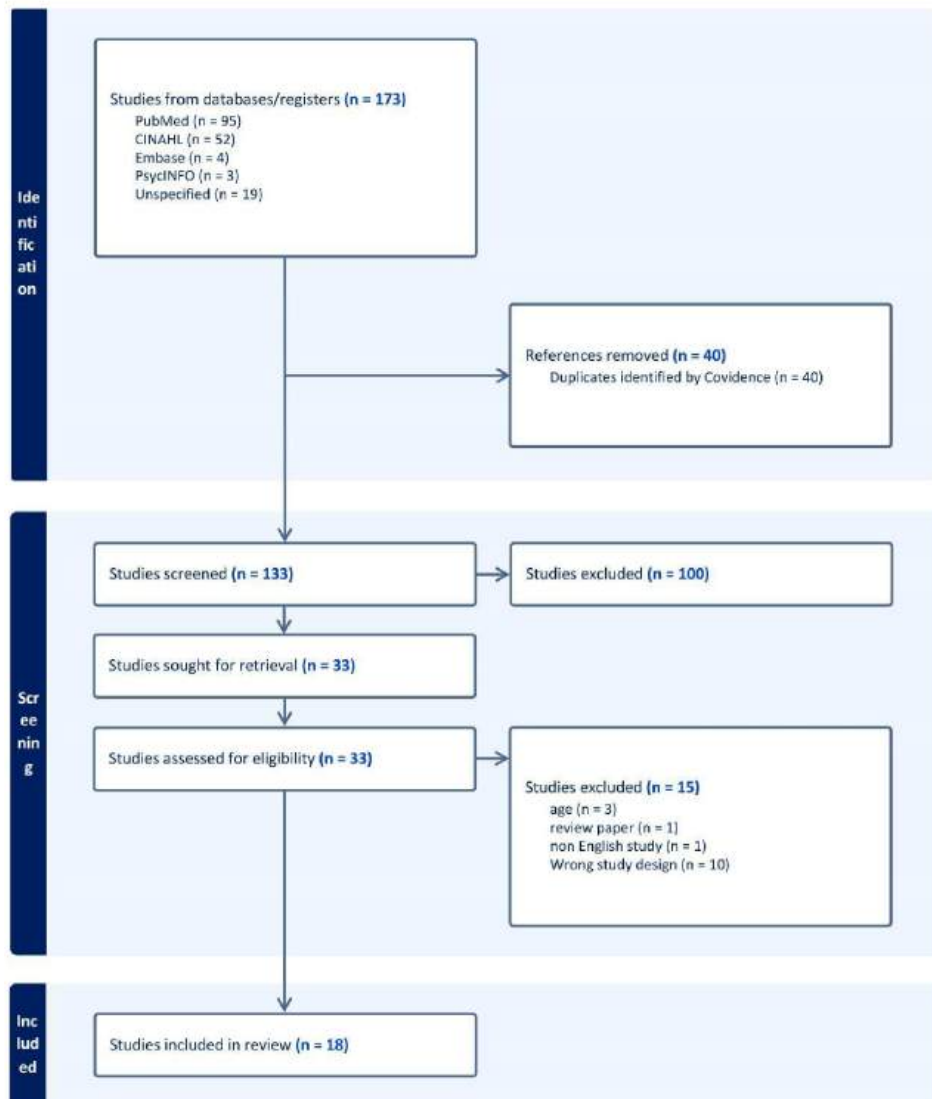
ADHD is diagnosed among 1 in 20 children worldwide (Faraone et al., 2015). Children with ADHD may also have an increased vulnerability to the rapid feedback and high-stimulation nature of video games, which has been reported in previous research, revealing that during video game use in individuals with ADHD, there is an increased striatal dopamine release (which occurs when dopamine neurons undergo "burst firing") and activation of the brain's reward pathway which can lead to impaired responses toward partial schedules of reinforcement (Koepp et al., 1998). A cross-sectional study found that an increase in the duration of video game use is directly correlated with an increase in the severity of inattentiveness (Alrahili et al., 2023). That is, children with ADHD, especially those with greater impulsivity, may be more prone to developing a gaming disorder.

This systematic review aims to evaluate the longitudinal associations between video game use and ADHD among children and adolescents.

Methods

The preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines were used to ensure transparency and completeness in documenting the articles found (Figure 1)

The Relationship Between Video Game Use and Attention-Hyperactivity Disorder (ADHD) Among Children and Adolescents: A Systematic Review



(Page et al., 2020)

Figure 1. PRISMA chart of bibliography search pictured above.

Due to the recent increase in overall video game availability and use among children and adolescents, the review focuses on literature published from 2013 through 2023. There were several inclusion criteria. First, only studies with samples of children and adolescents (≤ 18 years old) were included. Next, the article study design had to be observational (case-control study, cohort, or cross-sectional study) due to applicability to real-world settings. The exclusion criteria for selected articles included review articles, articles on the adult population, articles not published in English, and all interventional studies.

The searched databases include PubMed, CINAHL, Embase, ERIC (EBSCO), ERIC (ProQuest), and PsycINFO. The search terms were *children*, *adolescents*, *video games*, and

ADHD. The last day of the search was 4 December 2023. The search resulted in 173 articles being identified. Covidence (a screening and extraction tool promoting blinded screening) was used throughout the screening process.

Following the initial search of the 173 articles, 40 duplicated articles were removed. The remaining 133 articles underwent a title and abstract screening, excluding 100 due to irrelevancy (wrong study design, adult studies, and non-English language). The remaining 33 articles underwent a full-text review. 18 articles were extracted from the full-text review for inclusion in the systematic review.

Results

As shown in Table 2, results were gathered from the 18 articles extracted. The results were listed based on nine categories: population description, description of studies, inclusion and exclusion criteria, measure of ADHD, measure of video games, and the association between video game use and ADHD.

Data Extraction

For each article, two independent investigators (OKA and KCA) extracted the lead author's name, study aims, population description, inclusion and exclusion criteria, total number of participants, evaluation of ADHD and video game use, and conclusions. Following the extraction by the two investigators, the Cohen's Kappa was 0.72, showing a substantial agreement between the reviewers (McHugh, 2012).

Table 1. The table below depicts the agreement between reviewers OKA and KCA.

Reviewer A	Reviewer B	A Yes B Yes	A Yes B No	A No B Yes	A No B No	Proportionate Agreement	Yes Probability	No Probability	Random Agreement	Cohen's Kappa
OKA	KCA	26	9	5	93	0.89474	0.06134	0.5651	0.62643	0.71822

Table 2. The table below lists the authors' names, the location of the study, the study design, the population description, the inclusion and exclusion criteria, the measure of ADHD and video game use, and the conclusions.

Author	Location	Study Design	Population Description	Inclusion criteria	Exclusion criteria	Measure of ADHD	Measure of Video Games	Conclusion
C.K. Raelin et al. (2018)	USA	Cohort study	3396 ninth-grade students, with a mean age of 15.5 years	1. Not having ADHD at baseline 2. Ninth-grade students	1. Parental consent was not provided. 2. Data was not offered during the fall 10th-grade assessment. 3. ADHD symptom-positive status at baseline. 4. Did not provide ADHD symptom data at any of the four follow-up times	1. Self-reported form measure of Diagnostic and Statistical Manual of Mental Disorder-4th Edition (DSM-IV)	1. Self-report Scale	There is a statistically significant but modest association between higher frequency of digital media use and subsequent symptoms of ADHD.
E. Bourchtein (2019)	USA	Longitudinal study	302 Adolescents (aged 12-14 years) in eighth grade, recruited from local public schools in Southeastern and Midwestern United States	1. Enrolled in eighth grade 2. Full-scale IQ ≥ 80 based on WASI-II 3. Enrolled in regular education classes.	1. Meeting diagnosis of ASD, bipolar disorder, dissociative or psychotic disorder 2. Previous diagnosis of organic sleep disorder (obstructive sleep apnea, narcolepsy, restless leg syndrome, periodic limb movement disorder) 3. Not meeting the criteria of having ADHD	1. The Children's Interview for Psychiatric Syndromes Parent Version (P-ChIPS)	1. Self-report Questionnaire	Adolescents with ADHD reported significantly greater video game use.
Nader Alrahili et al. (2023)	Saudi Arabia	Cross-Sectional Study	393 children and adolescents aged 12-16 years, 58.3% male vs. females, 41.7%	1. Children and adolescents aged 12-16 years	1. Residing outside Saudi Arabia 2. Non-Arabic-speaking children and adolescents 3. Non-Arabic speaking guardians. 4. Intellectual disability	1. Arabic version of the ADHD Rating Scale-IV (ADHD-RS)	1. Game Addiction Scale (GAS-7)	Excessive video game playing negatively influences mental health and leads to issues including anxiety, depression, and ADHD
Hilario Blasco-Fontecilla et al. (2023)	Spain	Cross-sectional Study	51 children and adolescents diagnosed with ADHD between 7 and 17 years	1. 51 children and adolescents diagnosed with ADHD between 7 and 17 years	1. Mental retardation 2. Generalized developmental disorders 3. Neurological or psychiatric alterations that could compromise cognition	1. Spanish version of the Swanson, Nolan, and Pelham Scale for parents (SNAP-IV) 2. Abbreviated Conners Rating Scale for Parents (CPRS-HI)	1. ADITEC Video Games questionnaire (ADITEC-V)	Patient with Internet Game Disorder (IGD) tends to suffer from a predominantly hyperactive ADHD subtype in a more significant proportion than patients without IGD
Christopher J. Ferguson et al. (2022)	Singapore	Cohort study	3034 students from 6 primary schools and six secondary schools in Singapore, 72.8% reported being male	1. All students within the general education program	1. Parent refusal to consent	1. Self-report scale used to measure ADHD. (DuPaul, Power, Anastopoulos, & Reid, 1998)	1. Self-report scale	Youth with ADHD were more likely to develop later pathological gaming problems, while pathological gaming is not predictive of later mental health problems
Christopher J. Ferguson et al. (2021)	Singapore	Longitudinal study	3034 youth from Singapore, 72.8% are male	1. None reported	1. None reported	1. Self-report measure ADHD. (DuPaul, Power, Anastopoulos, & Reid, 1998)	1. Entertainment Software Ratings Board (ESRB)	Aggressive Video Games are not a risk factor for later mental health symptoms, including ADHD
Hyunsul Jeong et al. (2021)	South Korea	Cohort study	2319 3rd, 4th, and 7th graders	1. Presence in the 3rd, 4th, or 7th grades	1. Intellectual Disability of the Adolescent or a lack of competence in the Korean language in the parent or guardian.	1. A parent-report version of the Korean version of the ADHD rating scale (K-ARS)	1. Internet Game Use-Elicited Symptom Screen, a self-reported questionnaire based on the DSM-5 IGD criteria.	Among adolescents in South Korea, the existence of ADHD symptoms and spending more than 4 hours per day playing online games were associated with the occurrence or persistence of high

									risk for internet gaming disorder
Xuedi L et al. (2021)	Canada	Cohort Study	2026 children and youth aged 6-18 years in Ontario, Canada.	1. None reported	1. None reported	1. Strengths and Difficulties Questionnaire (SDQ) 2. Strengths and Weakness of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale (SWAN)	1. Self-report questionnaire	There was sufficient evidence to conclude the associations with symptoms of irritability, inattention, hyperactivity, and hyperactivity/impulsivity, with the effect sizes of these associations increasing with higher video game time.	
Micah Mazurek et al. (2013)	United States	Cross-Sectional Study	Parents of 141 boys aged 8-18 with ASD, ADHD, or Typical Development (TD)	1. None Reported	1. None Reported	1. Vanderbilt Attention Deficit/Hyperactivity Disorder Parent Rating Scale (VADPRS)	1. Self-report questionnaire	Children with ADHD may be at a higher risk for developing significant problems related to video game play, including excessive and problematic video game use. Attention problems in particular, are associated with problematic video game play for children with ADHD.	
Linda Pagani et al. (2022)	Canada	Cohort Study	1327 children and their families	1. None Reported	1. Being Indigenous or foreign nationals	1. Children self-reported their psychosocial characteristics using the Social Behaviour Questionnaire	1. Self-Report Scale	For both boys and girls, a more intense video gaming lifestyle predicted subsequent risks for reactive aggression and ADHD symptoms, compared with their same-sex counterparts reporting less intensity.	
J. L. Parraga et al. (2019)	Spain	Cross-Sectional Study	A total of 160 children and adolescents aged between 6 and 16 years participated in the study. All participants were Caucasian and lived in the Comunidad de Madrid. All of them belonged to high socioeconomic status according to Graffar's scale	1. Medication-naïve patients at the time of study	1. Neurodevelopmental disorders, including intellectual disability, autism spectrum disorder, Learning and language-related disorders, Evidence of neurological disorders, and History of psychiatric disorders such as psychosis and bipolar disorder.	1. Clinical diagnosis of ADHD according to DSM-IV-TR criteria	1. Self-Report Questionnaire	Results suggest that patients with ADHD spent more time than control playing video games and watching TV	
Katayoon Razjouyani et al. (2020)	Iran	Case-Control Study	99 children and adolescents with ADHD and 99 children and adolescents without the disorder in elementary school in Tehran	1. None Reported	1. None Reported	1. Conner's scale	1. Young's Internet Addiction Test	Addiction to video games was higher among children with ADHD than healthy controls.	
Lan Shuai et al. (2021)	China	Cross-sectional study	192 participants aged 8-16 years who met the	1. All participants met the diagnostic	1. Major sensory-motor disorders History of brain	1. Swanson, Nolan, Pelham Rating Scale (SNAP)	1. Young's Internet	The ADHD children with problematic digital media use (PDMU) had	

			diagnostic criteria for ADHD.	requirements for ADHD	damage Epilepsy Diagnosis of autism spectrum disorder		Addiction Test (YIAT)	significantly worse symptoms for inattention
Frode Stenseng et al.(2020)	Norway	Cross-Sectional Study	905 Norwegian speaking children from ages 6-10	1. None reported	1. None Reported	1. Preschool Age Psychiatric Assessment/Child and Adolescent Psychiatric Assessment (PAPA/CAPA)	1. Self-report questionnaires	Digital gaming did not predict more ADHD symptoms at any age, but more ADHD symptoms at age 8 predicted more gaming at age 10 years
Gabriel Arantes Tiraboschi et al.(2022)	Canada	Cohort Study	1467 Canadian youth aged 12-13 years.	1. Only families that speak English or French and whose pregnancy lasted between 24 and 42 weeks were included.	1. None Reported	1. Child Behavior Checklist and Preschool Behavior Questionnaire - self-report	1. Self-report questionnaire	Playing video games can represent a risk factor for developing attention problems in early adolescence.
Sofia Vadlin et al. (2016)	Sweden	Cohort Study	1868 Adolescents aged 12-16 (mean age = 15.39, median age = 15) were born in 1997 and 1999 and live in Västmanland, Sweden.	1. None Reported	1. Moving out of Vastmanland County Mental disabilities severe illness language difficulties Live in Sweden for less than five years	1. The Adult ADHD Self-Report Scale Adolescent version (ASRS-A)	1. The Gaming Addiction Identification Test (GAIT)	Having symptoms of ADHD increases the probability of concurrent symptoms of problematic gaming.
Jasmina Wallace et al. (2023)	Canada	Cohort Study	In the Greater Montreal Area, 3779 adolescents, 49 % girls, and the mean age is 12.8 years.	1. None Reported	1. None Reported	1. Strength and Difficulties Questionnaire (SDQ)	1. Self-report questionnaire	Increases in video game use were associated with a worsening of ADHD symptoms.
Lutz Wartberg et al. (2019)	Germany	Cross-Sectional Study	1095 family dyads and adolescents aged 12-14 years and a related caregiver)	1. None Reported	1. None Reported	1. Strengths and Difficulties Questionnaire (SDQ)	1. Internet Gaming Disorder Scale (IGDS)	Hyperactivity/inattention plays an essential role in the development of Internet Game Disorder (IGD)

Population description

The population description ranged across different regions and participant groups studied. Seven of the reviewed articles were conducted in North America (Ra et al., 2018; Burchtein et al., 2019; Li et al., 2021; Pagani et al., 2022; Wallace et al., 2023; Mazurek & Engelhardt, 2013; Tiraboschi et al., 2022), six were conducted in Asia (Alrahili et al., 2023; Ferguson et al., 2022; Ferguson & Wang, 2021; Jeong et al., 2021; Shuai et al., 2021; Khademi et al., 2020), and the rest of the reviewed articles were conducted in Europe (Wartberg et al., 2018; Blasco-Fontecilla et al., 2023; Párraga et al., 2019; Stenseng et al., 2019; Vadlin et al., 2016). Two papers included the same cohort of participants (Ferguson et al., 2022; Ferguson & Wang, 2021). The sample size of the studied population ranged from 51 participants to 3,779 participants. Seven reviewed papers included a participant count of over 2,000 people (Ferguson et al., 2022; Ferguson & Wang, 2021; Ra et al., 2018; Jeong et al., 2021; Li et al., 2021; Pagani et al., 2022; Wallace et al., 2023). Three studies included responses from family members of the studied samples (Pagani et

al.,2022; Mazurek & Engelhardt, 2013; Wartberg et al.,2018). The age ranges between the studies varied, with the population range falling between 6 to 18 years of age.

Description of Methodology

The methods of studies conducted were diverse. Eight of the reviewed papers were cohort studies (Ferguson et al., 2022; Ra et al., 2018; Jeong et al., 2021; Li et al., 2021; Pagani et al.,2022; Wallace et al., 2023; Tiraboschi et al., 2022; Vadlin et al., 2016), seven were cross-sectional studies (Alrahili et al., 2023; Mazurek & Engelhardt, 2013; Wartberg et al.,2018; Shuai et al., 2021; Blasco-Fontecilla et al.,2023; Párraga et al.,2019; Stenseng et al.,2019), two were longitudinal studies (Ferguson & Wang, 2021; Bourchtein et al., 2019), and one was a case-control study (Khademi et al.,2020). The case-control study examined the frequency of video game addiction in children with attention-deficit hyperactivity disorder among 99 children and adolescents with ADHD and 99 typical children and adolescents, revealing that addiction to video games was higher among children with ADHD than children with normative development.

Inclusion/Exclusion Criteria

The inclusion and exclusion criteria varied across the studies examined. Nine reviewed articles reported no inclusion criteria (Ferguson & Wang, 2021; Li et al., 2021; Pagani et al.,2022; Wallace et al., 2023; Mazurek & Engelhardt, 2013; Wartberg et al.,2018; Khademi et al.,2020; Stenseng et al.,2019; Vadlin et al., 2016), while eight reported no exclusion criteria (Ferguson & Wang, 2021; Li et al., 2021; Wallace et al., 2023; Mazurek & Engelhardt, 2013; Wartberg et al.,2018; Tiraboschi et al., 2022; Khademi et al.,2020; Stenseng et al.,2019). Only families that speak English or French were reported as an inclusion criterion in another paper (Tiraboschi et al., 2022). One reviewed paper reported non-Arabic speaking children, adolescents, guardians, and residents outside of Saudi Arabia as an exclusion criterion (Alrahili et al., 2023). Another paper reported moving out of Vastmanland County, Sweden, as an exclusion criterion (Stenseng et al.,2019). Neurodevelopmental disorders, including Autism Spectrum Disorder (ASD), intellectual disability, and learning disorders, were identified as exclusion criteria in eight of the reviewed articles (Alrahili et al., 2023; Ra et al., 2018; Bourchtein et al., 2019; Jeong et al., 2021; Shuai et al., 2021; Blasco-Fontecilla et al.,2023; Párraga et al.,2019; Vadlin et al., 2016). Two of the reviewed articles included the diagnosis of bipolar disorder, dissociative, or psychotic disorder as an exclusion criterion (Bourchtein et al., 2019; Párraga et al.,2019)

Measure of ADHD

Numerous scales were used to measure ADHD. Four reviewed articles used a self-report scale to evaluate ADHD (Ra et al., 2018; Pagani et al.,2022; Tiraboschi et al., 2022; Vadlin et al., 2016) and measured symptoms of impulsivity, hyperactivity, and inattentiveness based on DSM-IV and DSM-V criteria. Other reported scales used to assess ADHD were the Vanderbilt Attention Deficit/Hyperactivity Disorder Parent Rating Scale (Mazurek & Engelhardt, 2013), which identifies inattentive, hyperactive, or combined subtypes of ADHD; it contains an

eight-item performance and social functioning subscale. The Conners ADHD rating scale (Khademi et al.,2020; Blasco-Fontecilla et al.,2023) measures behavioral problems linked to ADHD, such as hyperactivity, aggressive behavior, and potential for violence. The Swanson, Nolan, Pelham Rating Scale rating scale (SNAP) (Shuai et al., 2021; Blasco-Fontecilla et al.,2023) measures the core symptoms of ADHD. The Children's Interview for Psychiatric Syndromes-Parent version (P-ChIPS) (Bourchtein et al., 2019) screens for ADHD and psychosocial stressors. The Strengths and Weakness of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale (SWAN) (Li et al., 2021) is a parent or teacher rating scale used to assess signs of ADHD in children. The Strengths and Difficulties Questionnaire (SDQ) (Li et al., 2021; Wallace et al., 2023; Wartberg et al.,2018) assesses children and adolescent's behavioral and emotional problems.

Measure of Video Game

Various scales were used to evaluate video game use in the reviewed articles. These scales include the Game Addiction Scale (GAS-7) (Alrahili et al., 2023), the Young's Internet Addiction Test (Shuai et al., 2021; Khademi et al.,2020), the Internet Gaming Disorder Scale (IGDS) (Wartberg et al.,2018), and the ADITEC Video Games (ADITEC-V) questionnaire (Blasco-Fontecilla et al.,2023). Other reviewed articles used either a self-report scale or a questionnaire to report the use of video games (Ferguson et al., 2022; Ra et al., 2018; Bourchtein et al., 2019; Jeong et al., 2021; Li et al., 2021; Pagani et al.,2022; Wallace et al., 2023; Mazurek & Engelhardt, 2013; Tiraboschi et al., 2022; Párraga et al.,2019; Stenseng et al.,2019). The self-report scales measured video game use based on the frequency and severity of video game use. The ADITEC-V questionnaire provides information regarding obsessive gambling, tolerance abstinence, and interference with activities (Khademi et al.,2020). The GAS-7 scale is a 7-item gaming addiction scale based on DSM criteria to evaluate gaming addiction. The Young's Internet Addiction Test measures symptoms associated with excessive internet use, while the IGDS measures the severity of gaming addiction. The self-report scales predominantly measured the frequency of video game use and the severity of symptoms following video game use.

The Association Between Video Game Use and ADHD

There were variable findings regarding the associations between video game use and ADHD across the majority of papers examined. Eight of the reviewed articles reported that children/adolescents with ADHD tend to have significantly more video game use and are more likely to develop pathological or excessive gaming problems (Ferguson et al., 2022; Bourchtein et al., 2019; Jeong et al., 2021; Mazurek & Engelhardt, 2013; Wartberg et al.,2018; Khademi et al.,2020; Párraga et al.,2019; Vadlin et al., 2016). A paper by Mazurek et al. (2013) further clarified that an inattentive type of ADHD is associated with problematic video game play for children with ADHD (Wallace et al., 2023). Likewise, six of the reviewed articles reported that video game use leads to an increased risk of developing ADHD (Alrahili et al., 2023; Ra et al., 2018; Pagani et al.,2022; Wallace et al., 2023; Tiraboschi et al., 2022; Blasco-Fontecilla et

al.,2023). Two of the reviewed articles found that digital gaming or aggressive video game use are not risk factors for developing ADHD (Ferguson & Wang, 2021; Stenseng et al.,2019). One paper reported that children with IGD tend to suffer from hyperactive ADHD (Ferguson & Wang, 2021). Another paper assessed the duration of time spent playing video games and their effects on the severity of IGD and ADHD and found that children and adolescents exhibiting ADHD symptoms who spent over 4 hours per day playing online games were at a higher risk of developing internet gaming disorder than children (AlYousefi et al.,2021). Similarly, Li et al. (2021) evaluated the associations between irritability, inattention, and hyperactivity symptoms with higher video game time. This relationship persisted after post hoc analysis between May 2020 and April 2021.

Discussion

Parents, teachers, clinicians, and policymakers have been greatly concerned about the use of video games among children and adolescents due to their effect on mental health. Therefore, this review, which had variable findings, was important to conduct on the longitudinal associations between video game use and ADHD in children and adolescents.

Confounding factors like family perceptions, teachers' perceptions, intelligence quotient (IQ), video game genres, and participants' genders make it challenging to assess video games' impact on cognition, similar to previous literature findings (Strahler Rivero et al., 2015). Another factor documented at an earlier literature that accounts for the variability of results is the lack of uniformity in the scales used to measure video game use and ADHD (Kaye et al.,2020). Most scales used to determine video game use were self-report questionnaires; as a result, there are inconsistencies in the information gathered. Another cause of the variability is differences in what is quantified as excessive video game use in the questionnaires. In addition, a review article by Polanczyk et al. (2007) reported that using informants (teachers or parents) led to higher reported prevalence rates than using diagnostic tests (Larsson et al.,2013).

Previous literature has documented greater problematic video game use among children with ADHD in comparison to children without ADHD (Kietglaiwansiri & Chonchaiya, 2018; Baer et al.,2012). Others reported that children with ADHD are addicted to the internet, including video games, more than children without ADHD (Weinstein et al.,2015). In addition, another paper found significant associations between the severity of ADHD symptoms and the severity of video games associations (Yoo et al., 2004). These findings are in line with two of the reviewed articles that found that children with ADHD were more likely to develop severe video game problems and a gaming disorder than children without ADHD (Jeong et al., 2021; Mazurek & Engelhardt, 2013).

One of the most common findings in this review was the worsening of ADHD symptoms, particularly inattentiveness and impulsivity following video game use. This is in line with previous findings, which revealed that gaming correlates with impulsivity, attentional problems, and poor self-control (Swing et al., 2010; Gentile et al., 2012). Another review article found a significant correlation between screen time (including video games) and externalizing behavior

problems in Middle East participants (Eirich et al.,2022). In addition, another article found that adolescents who play more than one hour of video games may have more intense symptoms of ADHD or inattention than those who do not have ADHD (Chan & Rabinowitz, 2006).

A small minority of papers found that digital gaming or aggressive video game use are not risk factors for developing ADHD (Ferguson & Wang, 2021; Stenseng et al.,2019). This unique finding was explained by studies that suggested that aggressive video game use is consistent with standard adolescent development (Olson, 2010) and that society's view on the impact of video games on mental health is exaggerated, which was conceptualized as moral panic theory (Kowert & Quandt, 2016).

Few of the reviewed articles attempted to establish a temporal relationship between gaming activities and ADHD symptoms. They found that an increase in video game use was associated with a worsening of ADHD symptoms, which is similar to other studies suggesting that excessive screen time in early childhood is related to attentional difficulties as time progresses (Christakis et al.,2004; Swing et al.,2010; Panagiotidi, 2017; Chan & Rabinowitz, 2006; Gentile et al., 2011).

Few of the reviewed articles reported on the subtypes of ADHD subtypes (impulsivity, hyperactivity, inattention), and this observation was previously reported in a meta-analysis by Nikkelen et al. (2014). The meta-analysis found that few studies evaluated the relationship between video game use and inattentiveness, and most of the studies reviewed used composite measures of ADHD symptoms (Nikkelen et al., 2014). Other findings from this review reported that the hyperactive subtype of ADHD is predominantly found among younger children than older children because hyperactivity decreases with age (Galéra, 2011; Lahey et al.,1994; Lahey et al.2005), and it plays a significant role in the development of IGD. The relationship between ADHD and video game use has yielded inconsistent results. Some articles report improvement in cognition (attention span) following video game use, whereas others report more susceptibility to video game addiction among patients with ADHD. Previously reported evidence suggested that children and adolescents diagnosed with ADHD, mostly the impulsivity trait, have a higher risk of being addicted to video games due to cortical hypoactivity (Romo et al.,2014; Yen et al., 2017). Cortical hypoactivity is related to a lack of control of impulses, time management, and greater sensitivity to sounds, lights, and immediate rewards (Makris et al., 2009).

In contrast, several neurobiological explanations suggest that video game play is beneficial and may improve attention span by activating neural networks by training distractor suppression mechanisms and aiding in reading (Hedges et al.,2013). Other findings suggest that video games activate the dorsolateral prefrontal region and insular subregions and improve connectivity between the temporoparietal and frontal lobes (Bavelier & Green et al.,2019; Antzaka et al., 2017; Gong et al., 2015; Föcker et al., 2019). These biological reasons explain the findings in some studies showing improvement in attention span after playing action video games (Steenbergen et al., 2015; Gan et al., 2020; Yao et al., 2020; Kollins et al., 2020) and the improvement in cognition among children and adolescents who play video games.

There are conflicting reports regarding the prevalence of ADHD in children and adolescents and video game usage rates in North America and Europe. Video game addiction ranged between 1 and 15% in Europe, 3 and 8.5% in the U.S., 14% in the U.K. and Korea, and 17% in Iran (Rajab et al., 2020; Saquib et al., 2017; AlYousefi et al., 2021; Jeong et al., 2021). In contrast, other researchers found that the prevalence rates of ADHD in Europe are much lower than in North America (Bird, 1998; Anderson, 1996; Timimi & Taylor, 2004). A meta-analysis found that the prevalence of ADHD does not differ significantly between countries in Europe, Africa, Australia, Asia, and the Americas (Polanczyk et al., 2007). However, this systematic review noted a lack of studies from Africa, South America, and the Middle East. This scarcity calls for more study on the cultural influence on relationships between video game usage and ADHD.

Limitations

This systematic review has potential limitations. First, the review did not include interventional studies, examined studies published from 2013 to 2023 due to the recent increased use of video games among children and adolescents, and excluded adult studies.

Despite these limitations, the systematic review suggests that children with ADHD are at a higher risk of developing a gaming disorder, and video games can exacerbate ADHD symptoms. Based on the data, more studies must be carried out to expand knowledge about the impact video game use has on children with ADHD and the causal relationship between video game use and ADHD.

Conclusion

This systematic review focused on the longitudinal associations between Attention-Deficit/Hyperactivity Disorder and video game use among children and adolescents. Although the findings were variable, the most common finding was that children with ADHD tend to have significantly more video game use and are more likely to develop problematic gaming usage.

A uniform rating scale for the measure of ADHD and video game usage should be designed to reduce the number of inconsistencies in the data obtained from studies evaluating the relationship between video game use and ADHD due to the vast number of rating scales. Responses from family members and teachers should be included in self-report scales to substantiate the responses from children and adolescents. Finally, more research should be conducted in underrepresented areas such as South America, the Middle East, and Africa to study the relationships between video game use and ADHD, the factors affecting these relationships, and to understand better the global patterns of video game addiction among youth.

Acknowledgments

I want to thank my mentor, Catherine Gonzalez, from Vanderbilt University, for her guidance in writing this research paper. I would also like to thank the Polygence research program for providing me with the means and resources to write this research paper.

Works Cited

- Alrahili, N., et al. "The Prevalence of Video Game Addiction and Its Relation to Anxiety, Depression, and Attention Deficit Hyperactivity Disorder (ADHD) in Children and Adolescents in Saudi Arabia: A Cross-Sectional Study." *Cureus*, vol. 15, no. 8, 4 Aug. 2023. Available from: <https://www.cureus.com/articles/173219-the-prevalence-of-video-game-addiction-and-its-relation-to-anxiety-depression-and-attention-deficit-hyperactivity-disorder-adhd-in-children-and-adolescents-in-saudi-arabia-a-cross-sectional-study#>
- AlYousefi, N. A., et al. "Exploring the Predictors of Depression Among Saudi Adolescents: Time for Urgent Firm Actions." *Saudi Medical Journal*, vol. 42, no. 6, 1 June 2021, pp. 673–81. Available from: <https://pubmed.ncbi.nlm.nih.gov/34078731/>
- Anderson, J. C. "Is Childhood Hyperactivity the Product of Western Culture?" *The Lancet*, vol. 348, no. 9020, July 1996, pp. 73–74.
- Antzaka, A., et al. "Enhancing Reading Performance through Action Video Games: The Role of Visual Attention Span." *Scientific Reports*, vol. 7, no. 1, 6 Nov. 2017. Available from: <https://www.nature.com/articles/s41598-017-15119-9>
- Banerjee, T. D., Middleton, F., and Faraone, S. V. "Environmental Risk Factors for Attention-Deficit Hyperactivity Disorder." *Acta Paediatrica*, vol. 96, no. 9, Sept. 2007, pp. 1269–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/17718779/>
- Baer, S., et al. "Electronic Media Use and Addiction among Youth in Psychiatric Clinic versus School Populations." *The Canadian Journal of Psychiatry*, vol. 57, no. 12, Dec. 2012, pp. 728-35
- Bavelier, D., and Green, C. S. "Enhancing Attentional Control: Lessons from Action Video Games." *Neuron*, vol. 104, no. 1, Oct. 2019, pp. 147–63.
- Beck, S. J., et al. "A Controlled Trial of Working Memory Training for Children and Adolescents with ADHD." *Journal of Clinical Child & Adolescent Psychology*, vol. 39, no. 6, 11 Nov. 2010, pp. 825-36
- Bird, H. R. "The Prevalence and Cross-Cultural Validity of Attention Deficit Hyperactivity Disorder." NIH Consensus Development Conference on Diagnosis and Treatment of Attention Deficit Hyperactivity Disorder, 1998.
- Blasco-Fontecilla, H., et al. "Lack of Educational Impact of Video Game Addiction in Children and Adolescents Diagnosed with ADHD: A Cross-Sectional Study." *Frontiers in Psychiatry*, 20 Apr. 2023, vol. 14.
- Bourchtein, E., et al. "Technology Use and Sleep in Adolescents With and Without Attention-Deficit/Hyperactivity Disorder." *Journal of Pediatric Psychology*, vol. 44, no. 5, 4 Jan. 2019. Available from: <https://academic.oup.com/jpepsy/advance-article-abstract/doi/10.1093/jpepsy/jsy101/5273623>
- Brown, Jane D, and Elizabeth M Witherspoon. "The Mass Media and American Adolescents' Health." *Journal of Adolescent Health*, vol. 31, no. 6, Supplement, 1 Dec. 2002, pp. 153–170, www.sciencedirect.com/science/article/pii/S1054139X02005074?casa_token=W5dRimvhc0AAAAA:4yD0vplxxnswtUmN4FjHrdhYLxIyeYrKtlkkLPzqxaO8FoPijyk_F53VEm_X16ygDNrkAX-4w#, [https://doi.org/10.1016/S1054-139X\(02\)00507-4](https://doi.org/10.1016/S1054-139X(02)00507-4).

- Chan, P. A., and Rabinowitz, T. "A Cross-Sectional Analysis of Video Games and Attention Deficit Hyperactivity Disorder Symptoms in Adolescents." *Annals of General Psychiatry*, vol. 5, no. 1, 24 Oct. 2006.
- Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., and McCarty, C. A. "Early Television Exposure and Subsequent Attentional Problems in Children." *Pediatrics*, vol. 113, no. 4, 1 Apr. 2004, pp. 708–13.
- Eirich, R., et al. "Association of Screen Time with Internalizing and Externalizing Behavior Problems in Children 12 Years or Younger." *JAMA Psychiatry*, vol. 79, no. 5, 16 Mar. 2022
- Faraone, S. V., et al. "Attention-Deficit/Hyperactivity Disorder." *Nature Reviews Disease Primers*, vol. 1, no. 15020, 6 Aug. 2015. Available from: <https://www.nature.com/articles/nrdp201520>
- Ferguson, C. J., Jeong, E. J., and Wang, J. C. K. "Pathological Gaming: A Longitudinal Study from the Perspectives of Mental Health Problems and Social Stress Model." *The Journal of General Psychology*, 8 Apr. 2022, pp. 1–21.
- Ferguson, C. J., and Wang, C. K. J. "Aggressive Video Games Are Not a Risk Factor for Mental Health Problems in Youth: A Longitudinal Study." *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 1, 1 Jan. 2021, pp. 70–3.
- Föcker, J., et al. "Neural Correlates of Enhanced Visual Attentional Control in Action Video Game Players: An Event-Related Potential Study." *Journal of Cognitive Neuroscience*, vol. 31, no. 3, Mar. 2019, pp. 377–89.
- Forsans E. De plus en plus de joueurs chez les 2-17 ans (étude NPD) [Internet]. AFJV. 2011 [cited 2024 May 15]. Available from: https://afjv.com/news/233_kids-and-gaming-2011.htm#:~:text=According%20to%20Kids%20and%20Gaming
- Galéra, C. "Early Risk Factors for Hyperactivity-Impulsivity and Inattention Trajectories from Age 17 Months to 8 Years." *Archives of General Psychiatry*, vol. 68, no. 12, 1 Dec. 2011, pp. 1267.
- Gan, X., et al. "Action Real-Time Strategy Gaming Experience Related to Increased Attentional Resources: An Attentional Blink Study." *Frontiers in Human Neuroscience*, vol. 14, 10 Apr. 2020.
- Gentile, D. A., et al. "Internet Gaming Disorder in Children and Adolescents." *Pediatrics*, vol. 140, Supplement 2, Nov. 2017, pp. S81–5. Available from: https://pediatrics.aappublications.org/content/140/Supplement_2/S81
- Gentile, D. A., et al. "Pathological Video Game Use Among Youths: A Two-Year Longitudinal Study." *Pediatrics*, vol. 127, no. 2, 17 Jan. 2011, pp. e319–29. Available from: <https://pediatrics.aappublications.org/content/127/2/e319>
- Gentile, D. A., et al. "Video Game Playing, Attention Problems, and Impulsiveness: Evidence of Bidirectional Causality." *Psychology of Popular Media Culture*, vol. 1, no. 1, Jan. 2012, pp. 62–70. Available from: <https://www.apa.org/pubs/journals/releases/ppm-1-1-62.pdf>
- Gong, D., et al. "Enhanced Functional Connectivity and Increased Gray Matter Volume of Insula Related to Action Video Game Playing." *Scientific Reports*, vol. 5, no. 1, 16 Apr. 2015. Available from : <https://www.nature.com/articles/srep09763>

- Hedges, J. H., et al. "Play, Attention, and Learning: How Do Play and Timing Shape the Development of Attention and Influence Classroom Learning?" *Annals of the New York Academy of Sciences*, vol. 1292, no. 1, 1 July 2013, pp. 1–20. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3842829/>
- Jeong H, Yim HW, Lee S, Lee HK, Potenza MN, Lee H. Factors associated with severity, incidence or persistence of internet gaming disorder in children and adolescents: a 2-year longitudinal study. *Addiction*. 2021 Jan 28;116(7).
- Kaye, L., et al. "The Conceptual and Methodological Mayhem of 'Screen Time.'" *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, 22 May 2020, p. 3661
- Khademi, M., et al. "An Investigation into the Frequency of Addiction to Video Games in Children with Attention-Deficit Hyperactivity Disorder." *Journal of Family Medicine and Primary Care*, vol. 9, no. 2, 2020, pp. 669.
- Kietglaiwansiri, T., and W. Chonchaiya. "Pattern of Video Game Use in Children with Attention-Deficit-Hyperactivity Disorder and Typical Development." *Pediatrics International*, vol. 60, no. 6, June 2018, pp. 523–28.
- Ko, C. H., et al. "The Associations Between Aggressive Behaviors and Internet Addiction and Online Activities in Adolescents." *Journal of Adolescent Health*, vol. 44, no. 6, June 2009, pp. 598-605
- Koepp, M. J., et al. "Evidence for Striatal Dopamine Release During a Video Game." *Nature*, vol. 393, no. 6682, May 1998, pp. 266–68. Available from: <https://www.nature.com/articles/30498>
- Kollins, S. H., et al. "A Novel Digital Intervention for Actively Reducing Severity of Pediatric ADHD (STARS-ADHD): A Randomised Controlled Trial." *The Lancet Digital Health*, vol. 2, no. 4, April 2020, pp. e168–78.
- Kowert, R., and T. Quandt. *The Video Game Debate: Unravelling the Physical, Social, and Psychological Effects of Digital Games*. Routledge, 2016.
- Lahey, B. B., et al. "DSM-IV Field Trials for Attention Deficit Hyperactivity Disorder in Children and Adolescents." *American Journal of Psychiatry*, vol. 151, no. 11, Nov. 1994, pp. 1673–85.
- Lahey, B. B., et al. "Instability of the DSM-IV Subtypes of ADHD From Preschool Through Elementary School." *Archives of General Psychiatry*, vol. 62, no. 8, Aug. 2005, pp. 896.
- Li, X., et al. "Screen Use and Mental Health Symptoms in Canadian Children and Youth During the COVID-19 Pandemic." *JAMA Network Open*, vol. 4, no. 12, 28 Dec. 2021, article e2140875. Available from: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2787454>
- Larsson, H., et al. "Family Income in Early Childhood and Subsequent Attention Deficit/Hyperactivity Disorder: A Quasi-Experimental Study." *Journal of Child Psychology and Psychiatry*, vol. 55, no. 5, 23 Sept. 2013, pp. 428–35. Available from: <https://acamh.onlinelibrary.wiley.com/doi/10.1111/jcpp.12140>
- Makris N, Biederman J, Monuteaux MC, Seidman LJ. Towards Conceptualizing a Neural Systems-Based Anatomy of Attention-Deficit/Hyperactivity Disorder. *Developmental Neuroscience*. 2009;31(1-2):36–49.
- Mark AE, Janssen I. Relationship between screen time and metabolic syndrome in adolescents. *Journal of Public Health [Internet]*. 2008 Apr 2;30(2):153–60. Available from: <https://academic.oup.com/jpubhealth/article/30/2/153/1542221>

- Masi L, Abadie P, Herba C, Emond M, Gingras MP, Amor LB. Video Games in ADHD and Non-ADHD Children: Modalities of Use and Association With ADHD Symptoms. *Frontiers in Pediatrics*. 2021 Mar 12;9(632272).
- Mazurek MO, Engelhardt CR. Video Game Use in Boys With Autism Spectrum Disorder, ADHD, or Typical Development. *PEDIATRICS*. 2013 Jul 29;132(2):260–6.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica* [Internet]. 2012 Jan 1 [cited 2024 May 21];276–82. Available from: <https://www.biochemia-medica.com/en/journal/22/3/10.11613/BM.2012.031>.
- Menendez-García A, Jiménez-Arroyo A, Rodrigo-Yanguas M, Marin-Vila M, Sánchez-Sánchez F, Roman-Riechmann E, et al. Adicción a Internet, videojuegos y teléfonos móviles en niños y adolescentes: un estudio de casos y controles. *Adicciones*. 2020 Dec 4;34(3).
- Michael T, Seymour PA. A Brief Comprehensive Review of the Interactions Between Attention-Deficit/Hyperactivity Disorder and Internet Gaming Disorder. *International Journal of Mental Health and Addiction*. 2023 Dec 11;
- Muzwagi AB, Motiwala FB, Manikkara G, Rizvi A, Varela MA, Rush AJ, et al. How Are Attention-deficit Hyperactivity and Internet Gaming Disorders Related in Children and Youth? *Journal of Psychiatric Practice*. 2021 Nov;27(6):439–47.
- National Institute of Mental Health. Attention-Deficit/Hyperactivity Disorder [Internet]. www.nimh.nih.gov. National Institute of Mental Health; 2023. Available from: <https://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd>
- Nikkelen SWC, Valkenburg PM, Huizinga M, Bushman BJ. Media use and ADHD-related behaviors in children and adolescents: A meta-analysis. *Developmental Psychology*. 2014;50(9):2228–41.
- Olson CK. Children’s motivations for video game play in the context of normal development. *Review of General Psychology* [Internet]. 2010;14(2):180–7. Available from: http://ocw.metu.edu.tr/pluginfile.php/11519/mod_resource/content/1/gpr-14-2-180.pdf
- Pagani LS, Jalbert M, Derevensky JL. From one year to the next: Video gaming life-style predicts subsequent psychosocial risk in adolescent boys and girls. *Journal of Paediatrics and Child Health*. 2022 Jun 3;58(9):1554–60.
- Panagiotidi M. Problematic Video Game Play and ADHD Traits in an Adult Population. *Cyberpsychology, Behavior, and Social Networking*. 2017 May;20(5):292–5
- Párraga JL, Pérez BC, López-Martín S, Albert J, Fernández-Mayoralas DM, Fernández-Perrone AL, et al. Attention-deficit/hyperactivity disorder and lifestyle habits in children and adolescents. *Actas Españolas de Psiquiatría* [Internet]. 2019 Jul 1 [cited 2024 May 15];47(4):158–64. Available from: <https://actaspsiquiatria.es/index.php/actas/article/view/286>
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal* [Internet]. 2021 Mar 29;372(71). Available from: <https://www.bmj.com/content/372/bmj.n71>

- Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. *American Journal of Psychiatry* [Internet]. 2007 Jun 1;164(6):942–8. Available from: <https://ajp.psychiatryonline.org/doi/full/10.1176/ajp.2007.164.6.942>
- Ra, CK, et al. "Association of Digital Media Use with Subsequent Symptoms of Attention-Deficit/Hyperactivity Disorder Among Adolescents." *JAMA*, vol. 320, no. 3, 2018, pp. 255. <https://jamanetwork.com/journals/jama/fullarticle/2687861>.
- Rajab, AM, et al. "Gaming Addiction and Perceived Stress Among Saudi Adolescents." *Addictive Behaviors Reports*, vol. 11, June 2020, p. 100261.
- Romo, L, et al. "Gambling and Attention Deficit Hyperactivity Disorders (ADHD) in a Population of French Students." *Journal of Gambling Studies*, vol. 31, no. 4, Dec. 2014, pp. 1261–72.
- Saquib, N, et al. "Video Game Addiction and Psychological Distress Among Expatriate Adolescents in Saudi Arabia." *Addictive Behaviors Reports*, vol. 6, Dec. 2017, pp. 112-17
- Sampalo, M, et al. "Action Video Gaming and Attention in Young Adults: A Systematic Review." *Journal of Attention Disorders*, vol. 27, no. 5, Feb. 2023, pp. 530–38.
- Scassellati, C, et al. "Biomarkers and Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-Analyses." *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 51, no. 10, Oct. 2012, pp. 1003-1019.e20.
- Shuai, L, et al. "Influences of Digital Media Use on Children and Adolescents with ADHD During COVID-19 Pandemic." *Globalization and Health*, vol. 17, no. 1, Apr. 2021.
- Steenbergen, L, et al. "Action Video Gaming and Cognitive Control: Playing First Person Shooter Games Is Associated with Improved Action Cascading but Not Inhibition." *PLOS ONE*, edited by Weinstein AM, vol. 10, no. 12, Dec. 2015, p. e0144364.
- Stenseng, F, et al. "Time Spent Gaming and Psychiatric Symptoms in Childhood: Cross-Sectional Associations and Longitudinal Effects." *European Child & Adolescent Psychiatry*, Sept. 2019.
- Strahler Rivero, T, et al. "ADHD Rehabilitation through Video Gaming: A Systematic Review Using PRISMA Guidelines of the Current Findings and the Associated Risk of Bias." *Frontiers in Psychiatry*, vol. 6, Oct. 2015.
- Swing, EL, et al. "Television and Video Game Exposure and the Development of Attention Problems." *PEDIATRICS*, vol. 126, no. 2, July 2010, pp. 214–21. <https://pubmed.ncbi.nlm.nih.gov/20603258>.
- Timimi, S, and E Taylor. "ADHD Is Best Understood as a Cultural Construct." *British Journal of Psychiatry*, vol. 184, no. 01, Jan. 2004, pp. 8–9. <https://www.cambridge.org/core/journals/the-british-journal-of-psychiatry/article/adhd-is-best-understood-as-a-cultural-construct/6A2C79B3CD1D4D2627F3EE4C3DF3FBAF/core-reader>.
- Tiraboschi, GA, et al. "Associations Between Video Game Engagement and ADHD Symptoms in Early Adolescence." *Journal of Attention Disorders*, vol. 108705472110734, Jan. 2022
- Vadlin, S, et al. "Associations Between Problematic Gaming and Psychiatric Symptoms Among Adolescents in Two Samples." *Addictive Behaviors*, vol. 61, Oct. 2016, pp. 8–15.

- Wallace, Jasmina, et al. "Screen time, impulsivity, neuropsychological functions and their relationship to growth in adolescent attention-deficit/hyperactivity disorder symptoms." *Scientific Reports* 13.1 (2023): 18108.
- Wartberg, L, et al. "A Longitudinal Study on Psychosocial Causes and Consequences of Internet Gaming Disorder in Adolescence." *Psychological Medicine*, vol. 49, no. 2, Apr. 2018, pp. 287-94
- Weiss, MD, et al. "The Screens Culture: Impact on ADHD." *ADHD Attention Deficit and Hyperactivity Disorders*, vol. 3, no. 4, Sept. 2011, pp. 327–34.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3220824/>.
- Weinstein, A, et al. "Internet Addiction and Attention Deficit Hyperactivity Disorder (ADHD) Among Schoolchildren." *IMAJ | The Israel Medicine Association Journal*, vol. 17, no. 12, Dec. 2015
<https://www.ima.org.il/MedicineIMAJ/viewarticle.aspx?year=2015&month=12&page=731>.
- World Health Organization. "Gaming Disorder." www.who.int, 2023.
<https://www.who.int/standards/classifications/frequently-asked-questions/gaming-disorder>.
- World Health Organization. ICD-11. International Classification of Diseases 11th Revision, 2019. <https://icd.who.int/en>.
- Yao, Y, et al. "Action Real-Time Strategy Gaming Experience Related to Enhanced Capacity of Visual Working Memory." *Frontiers in Human Neuroscience*, vol. 14, Aug. 2020.
<https://dx.doi.org/10.3389/fnhum.2020.00333>.
- Yen, JY, et al. "Association Between Internet Gaming Disorder and Adult Attention Deficit and Hyperactivity Disorder and Their Correlates: Impulsivity and Hostility." *Addictive Behaviors*, vol. 64, Jan. 2017, pp. 308–13.
- YOO, HJ, et al. "Attention Deficit Hyperactivity Symptoms and Internet Addiction." *Psychiatry and Clinical Neurosciences*, vol. 58, no. 5, Oct. 2004, pp. 487–94.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1440-1819.2004.0129>

Justification of the Decision to Begin the First Crusade By Merrick Jiang

“Let those who have been robbers now become knights. Let those who have fought their brothers and relatives now fight the barbarians. Let those who have served as mercenaries for small pay now obtain the eternal reward. Let those who have worn themselves out in body and soul now work for double honor.” (“Deus Volt, God Wills It.”) This cry to war is what resounded the ears at Clermont, 1095. It was the word of Pope Urban II, the leader of the Roman Catholic church from 1088 to death in 1099. (Noble, Thomas F. X.) It called for the whole of Europe to abandon all internal oppositions to rise as one army against the enemy of the east. This undertaking came to be known as the First Crusade. But what kind of world could have led to such a predicament? The story leading up to the First Crusade can be picked up after the rise of the Seljuk Turks. The Seljuks were converted Muslims from central Asia who conquered and controlled the Middle East during the eleventh century. (Riley-Smith, Jonathan) Eventually, they turned westward to the Byzantine Empire, successfully seizing control of many Byzantine territories and ending up pushing all the way west until they were just a hundred miles from the Byzantine capital. (“Manzikert, August 1071.”) Finally, the Byzantine emperor reached out to Pope Urban II to the west for military assistance, which resulted in the declaration of the First Crusade at the Council of Clermont, 1095. (Riley-Smith, Jonathan) Despite the bloody connotations of a call to war, what many points of evidence argue is that certain circumstances that threatened Christendom justified Pope Urban II’s decision to commence the First Crusade. The main circumstances that will be discussed and supported are the strong implications of a return to Roman-like Christian persecution under Seljuk rule, the severe threat that the Seljuks imposed on all of Europe by endangering the Byzantine Empire, and lastly the hazardously divided continent of Europe which threatened the institution of the church.

First of all, it is indispensable to mention the preexistent terror that many Christians must have internalized in aversion to the memory of the torture that their forefathers suffered under Roman occupation. (Koussoumzin 50-53) For approximately two-hundred and fifty years, the first Christians were ruthlessly oppressed by the Roman Empire across numerous periods of varying intensity. (Koussoumzin 49-50) Various accounts of contemporary writers describe the horrors endured by believers; one writes, “If I had a hundred lips and a chest of iron I would still be unable to enumerate all the tortures suffered by the faithful,” (Koussoumzin 52) and another, “I myself witnessed how iron was blunted and broken, and the executioner tired out and had to be relieved by another.” (Koussoumzin 53) Examples of such persecution include execution of Christians for being Christian, torturing of Christians to give up the names of other Christians, torturing of Christians to renounce their Christianity, and the destruction of all Christian communities, churches, and books. (Koussoumzin 50-52) This presupposition of what foreign and especially non-Christian occupation implied must have been a key player in why the initialization of the First Crusade was justified to them. Their fighting would be justified as they fought to save their families and loved ones from the supposed torture to come. As one modern-day historian puts it, “Certainly the eleventh century was more receptive to a militia

Christi than to the old idea of athletes of Christ suffering torture.” (Hill) In short, one motivation of Urban that justified the declaration of the First Crusade must have been to act in self-defense against the Seljuks out of the fear of possible Roman-like persecution under Turkish occupation.

The claim that the Seljuk threat to Christianity justified Urban’s call to defense is further accentuated by the undeniability that the Seljuk Turks gravely endangered the existence of the Byzantine Empire. This endangerment is firstly made evident by the aftermath of the Battle of Manzikert, in which the Turkish forces completely encircled and destroyed the entire main Byzantine army. “... Romanus Diogenes then ordered the army to stand and fight. However, the commander of his rear guard, Andronicus Ducas (a rival of the emperor), ignored his orders and continued marching to the Byzantine camp. This betrayal allowed the Seljuks to surround and annihilate the main army.” (“Manzikert, August 1081.”) This crushing defeat ultimately backed the empire into a corner as it held off the Seljuks just a hundred miles away from Constantinople, its capital city. Although this endangerment of the Byzantine Empire is not a direct threat to all of Europe, Urban understood that Islamic conquest of the Byzantine Empire could likely lead to the conquest of Christian territories even further west. As one source puts it, “They [the west] knew that if Constantinople fell to the Muslim Seljuks, the empire probably would collapse entirely. They wanted the empire to remain stable and strong, for it served as a buffer between the Muslim empire and the Christian countries of Europe... With no Byzantine Empire to hold the Muslims in check, Europe would face an even greater threat. ” (“Origins of the Crusades.”) In other words, the Byzantine vulnerability to Seljuk conquest supports the narrative that the Seljuks threatened the survival of Christianity and obligated the west to intervene, because that vulnerability didn’t only threaten the Byzantines, but stood as a liability to the entire continent of Europe along with Christianity as a religion. Urban needed to protect the Byzantine empire in order to protect all of Christendom.

Finally, although the idea to protect Christians from Seljuk conquest may have been enough reason to initiate the First Crusade, there exists the third and last reason that a call for holy war would have unified Europe under one cause, saving Christianity from becoming any more divided. Throughout his entire rule as pope, “Urban faced continuous and stormy opposition from the German emperor, Henry IV. Throughout the pope's reign, he had to contend with antipopes who denied him full control of Rome.” (Noble, Thomas F. X.) To put it plainly, this circumstance of division and opposition threatened the survival of the church as it would have divided people away from what Urban believed to be the true branch of Christianity, the Roman Catholic Church. As one source states, “Pope Urban II had his own problems to deal with. Conflicts between the European kings and nobles were threatening the papacy and, he believed, Christendom itself. He needed a cause behind which to unite them.” (“The Crusades.”) In short, based on the circumstance that any more division of the church deeply threatened its survival, Urban was justified.

Nearing the close of the eleventh century, there existed a plethora of threats to the survival of Christianity which justified Urban’s call to arms at the Council of Clermont. Although war doesn’t ever get any less bloody, the blood-filled history that paved the origins of

the church paired with the existing implications of a possible return to such atrocity gave the Europeans the justified cause to protect Christianity from another period of persecution. This cause of self-defense was yet further amplified by the extreme state of vulnerability that the Byzantines faced, jeopardizing not only the safety of the Byzantine Empire alone but the safety of all of Christianity and Europe. Last but not least, the decision of the First Crusade was justified by the fact that the dire situation of the division created by Urban's enemies jeopardized the stability of Christianity and held the religion vulnerable to collapse. For words of finality, Urban's call of the First Crusade connects to a grand idea of triaging evil. There will always be times when there is no good decision, and where there will always be some left behind. But what is most important is to always roll with the punches, and to make the best of whatever must happen.

Works Cited

- "Deus Volt, God Wills It." Gale World History Online Collection, Gale, 2007. Gale In Context: World History, link.gale.com/apps/doc/BT2399700009/WHIC?u=mlln_m_brookhs&sid=bookmark-WHIC&xid=342f0016. Accessed 19 May 2024.
- Hill, John H. "Clermont, Council of." Dictionary of the Middle Ages, edited by Joseph R. Strayer, Charles Scribner's Sons, 1989. Gale In Context: World History, link.gale.com/apps/doc/BT2353200678/WHIC?u=mlln_m_brookhs&sid=bookmark-WHIC&xid=702690d0. Accessed 19 May 2024.
- Koulomzin, Sophie. "Persecutions." Heroes For Truth: The Orthodox Christian Church Through the Ages Volume I, Russian Orthodox Church of America, 1968, pp. 49-53.
- "Manzikert, August 1071." Gale Encyclopedia of World History: War, Gale, 2009. Gale In Context: Middle School, link.gale.com/apps/doc/EJ3048700044/MSIC?u=mlln_m_brookhs&sid=bookmark-MSIC&xid=5dda7d7e. Accessed 19 May 2024.
- Noble, Thomas F. X. "Urban II." World Book Advanced, 2024, www.worldbookonline.com/advanced/article?id=ar577741. Accessed 19 May 2024.
- Nordstrom, Justin. "The Pope." St. James Encyclopedia of Popular Culture Online, Gale, 2013. Gale In Context: Biography, link.gale.com/apps/doc/DIZRUY835406948/BIC?u=mlln_m_brookhs&sid=bookmark-BIC&xid=47381931. Accessed 19 May 2024.
- "Origins of the Crusades." The Crusades Reference Library, edited by Neil Schlager, et al., vol. 1: Almanac, UXL, 2005, pp. 48-68. Gale In Context: World History, link.gale.com/apps/doc/CX3441100013/WHIC?u=mlln_m_brookhs&sid=bookmark-WHIC&xid=3dec17. Accessed 19 May 2024.
- Riley-Smith, Jonathan. "Crusades." World Book Advanced, 2024, www.worldbookonline.com/advanced/article?id=ar142340. Accessed 19 May 2024.
- Russell, Mona L. "Seljuks." World Book Advanced, 2024, www.worldbookonline.com/advanced/article?id=ar500540. Accessed 19 May 2024.
- Shlosser, Franziska E. "Christians, Roman Persecution of." Encyclopedia of Genocide and Crimes Against Humanity, edited by Dinah L. Shelton, vol. 1, Macmillan Reference USA, 2005, pp. 177-179. Gale In Context: World History, link.gale.com/apps/doc/CX3434600071/WHIC?u=mlln_m_brookhs&sid=bookmark-WHIC&xid=e5e6e0d4. Accessed 19 May 2024.
- "The Crusades." Science and Its Times, edited by Neil Schlager and Josh Lauer, vol. 2, Gale, 2001. Gale In Context: Biography, link.gale.com/apps/doc/CV2643450125/BIC?u=mlln_m_brookhs&sid=bookmark-BIC&xid=41b032f6. Accessed 19 May 2024.
- "The Great Schism: 1054." Global Events: Milestone Events Throughout History, edited by Jennifer Stock, vol. 4: Europe, Gale, 2014. Gale In Context: World History, link.gale.com/apps/doc/FTQOAM844250807/WHIC?u=mlln_m_brookhs&sid=bookmark-WHIC&xid=f15599be. Accessed 19 May 2024.
- "Urban II." Historic World Leaders, edited by Anne Commire, Gale, 1994. Gale In Context: Biography, link.gale.com/apps/doc/K1616000588/BIC?u=mlln_m_brookhs&sid=bookmark-BIC&xid=6e2b6f76. Accessed 19 May 2024.

Chronic Kidney Disease in the Middle East By Michael Fam

Abstract

Chronic Kidney Disease (CKD) consists of five different stages, with the final stage, kidney failure, posing the greatest threat to those affected by it. CKD has many determinants, but the most prominent are diabetes and hypertension. Among the various forms of treatment that patients who possess CKD undergo are kidney transplantation and dialysis modalities, such as hemodialysis and peritoneal dialysis. Hemodialysis requires patients to visit dialysis centers multiple times a day while peritoneal dialysis can be performed at one's home. The accessibility and utilization of these kidney treatments in the Middle East is a growing concern. Although kidney transplants and dialysis modalities cost less in the Middle East than in North America, the Middle East's heavy reliance on kidney transplants from living donors and their underutilization of dialysis modalities shed light on the inferior patient outcomes in its regions than in developed nations, such as the United States.

Research Question

What are the determinants of kidney disease, and how do the accessibility and utilization of kidney transplantation and dialysis modalities, such as hemodialysis and peritoneal dialysis, impact patient outcomes in the Middle East?

Introduction

Chronic kidney disease refers to the decreased kidney function which has a glomerular filtration rate of less than 60 mL/min per 1.73 m² of around a duration of 3 months (Ammirati, 2020). Glomerular filtration rate (GFR) is the total fluid filtered through the functioning nephrons per unit of time. The criteria of CKD are that the GFR must be less than 60 and/or there must be 1 or more of the following: albuminuria, urinary sediment abnormality, electrolyte or other abnormality due to tubular disorder, abnormalities on histology, structural abnormalities detected by imaging, and history of kidney transplantation.

CKD is composed of five stages with each higher stage consisting of greater damage to the kidney: G1, G2, G3a, G3b, G4, and G5. In G1, the GFR is less than 90 (normal to high). In G2, the GFR is between 60 to 89 (mildly decreased). In G3a, the GFR is 45 to 59 (mildly to moderately decreased). In G3b, the GFR is 30 to 44 (moderately to severely decreased). In G4, the GFR is 15 to 29 (severely decreased). In G5, the GFR is less than 15 (kidney failure). The fifth and final stage of CKD is classified as kidney failure or end-stage kidney disease. Once this stage is reached, the damage to the kidney is greatest. On the list of causes of death, CKD rose from 13th in 2016 to 12th in 2017 and is projected to be 5th by 2024, highlighting the growing concern associated with the disease (Kovesdy, 2022). The ISN Middle East region consists of 13 countries: Bahrain, Iraq, the Islamic Republic of Iran, Jordan, Kuwait, Oman, Lebanon, Qatar, Saudi Arabia, the Syrian Arab Republic, United Arab Emirates, West Bank and Gaza, and Yemen (Karam et al., 2024). In 2023, the global median of CKD prevalence was 9.54% while the

Middle East median was 8.17%. Worldwide the death attributed to CKD is 2.44%, and in the Middle East, it is 3.36%. Additionally, worldwide the DALYs (the disability-adjusted year which combines lost years of life from premature mortality and lost years of life from living in regions that lack full health or lost years of life from a disability) is 1.47%, and in the Middle East it is 2.17%. Thus, the Middle East is more impacted by CKD compared to the global median.

While the causes of CKD vary from region to region, the main causes in both high-income and middle-income countries as well as the many low-income countries are diabetes and hypertension (Webster et al., 2017). Diabetes is the cause of 30 to 50% of CKD, affecting over 280 million adults across the world. In addition, those with worsening blood pressure control have an increased risk of developing CKD. In regions such as India, Asia, and sub-Saharan Africa, CKD can arise from glomerulonephritis and other unknown causes. Because of environmental pollution of water from heavy metals and of soil from organic compounds, there have been many localized epidemics of CKD in these regions. Aside from lifestyle and environmental problems, causes of CKD can be attributed to genetics and epigenetics, such as abnormalities of the kidney and urinary tract or autosomal dominant polycystic kidney disease. Injuries to the kidney tissues, such as renal fibrosis, can also lead to CKD. While there are steps to slow the progression of CKD, the disease is generally irreversible as any damage done to the kidney cannot be reversed. Despite being irreversible, there are some forms of treatment for CKD with kidney transplantation, hemodialysis, and peritoneal dialysis being the forms that will be discussed in this report.

Kidney transplantation is a surgery in which a healthy kidney is placed into a person who is suffering from kidney failure (Augustine, 2018). Since the 1990s, kidney transplants have provided a survival advantage to those who suffered from the end-stage renal disease over treatment based on dialysis. Moreover, the risk of death following a transplant is lower after a couple of months than it is for those on dialysis. There are two types of organ transplant: living-donor organ transplant and deceased-donor organ transplant. After one year, patient survival is greater than 90% for both groups. However, by five years, 80% of patients with a living-donor organ transplant survive compared to 65% of those with a deceased-organ transplant. Despite kidney transplantation serving as an optimal treatment method for those suffering from CKD, the growing demand for available kidneys is much greater than those that are available. For instance, in 2016, around 100,000 patients were on the kidney waiting list. However, only about 19,000 of those patients were able to receive one. In addition to kidney transplantation, there are also dialysis modalities that help treat people with kidney failure.

Among the dialysis modalities that are associated with CKD are hemodialysis and peritoneal dialysis. Hemodialysis is associated with the filtration of blood in a person who lacks working kidneys (Himmelfarb & Ikizler, 2010). More specifically, hemodialysis aims to restore the fluid environment of a functioning kidney by transporting solutes from the blood into the dialysate and solutes from the dialysate into the blood. Those who undergo hemodialysis are generally expected to go to dialysis centers two to three times a week with each session lasting approximately three to four hours (Zazzeroni et al., 2017). The other form of dialysis known as

peritoneal dialysis also filters blood inside the body (Andreoli & Totoli, 2020). However, peritoneal dialysis differs from hemodialysis in that it deals with the infusion of a sterile solution into the peritoneal cavity by use of a catheter. This form of dialysis also utilizes the peritoneal membrane as an exchange surface for solutes and water. Unlike hemodialysis, peritoneal dialysis can occur independently at one's home multiple times a day (Zazzeroni et al., 2017). Out of the 13 countries in the ISN Middle East region, 7 have access to transplant surgeons, 5 have access to transplant coordinators, 6 have access to dialysis nurses, and 4 have access to dialysis technicians (Karam et al., 2024). Thus, kidney transplantation and dialysis modalities in the ISN Middle East are not able to maximize patient outcomes for those suffering from CKD.

The aim of this report is to shed light on the accessibility of various forms of kidney treatment across the Middle East in comparison to the United States in order to ensure that those suffering from CKD can acquire proper care and have the greatest chance of survival. This report will cover more specifically the determinants of kidney failure, kidney transplants, and dialysis modalities in the Middle East to highlight the overall condition of CKD in these regions.

Determinants of Kidney Failure

The role of the kidney in the body is to maintain the constancy of the extracellular environment by removing the waste products of metabolism and adjusting urinal water and electrolyte excretion (Ferguson & Waikar, 2012). The kidney also regulates blood pressure, bone mineral metabolism, and red blood cell production. Healthy kidneys can filter about a half cup of blood every minute. The urine travels from each kidney to the bladder through ureters, which are a pair of thin tubes (*Causes of Chronic Kidney Disease - NIDDK*, n.d.). Diabetes is a major risk factor for CKD as it damages the kidneys. Those affected by diabetes have injured small blood vessels (*Diabetes - A Major Risk Factor for Kidney Disease*, 2015). Because the blood vessels in the kidneys are injured, the kidneys lose their ability to filter blood properly. As a result, the body can retain greater water and salt than it should as they cannot be excreted efficiently. One of the earliest signs of CKD from diabetes is protein in urine (*Causes of Chronic Kidney Disease - NIDDK*, n.d.). When the kidney's filtration system gets weakened, a protein known as albumin exits the blood and enters into the urine. Albumin is the main protein in the blood, and it is made by the liver. With a healthy kidney, the albumin remains in the blood and does not enter the urine. Diabetes may also damage nerves within the body which can lead to difficulty in emptying the bladder (*Diabetes - A Major Risk Factor for Kidney Disease*, 2015). The pressure from the full bladder could then back up and damage the kidneys. If urine remains in the bladder for a long period, an infection can arise from the rapid growth of bacteria within the urine where there is a high sugar level. Once the kidneys fail, the blood urea nitrogen (BUN) levels and the amount of creatinine in the blood will rise. Although having diabetes does not guarantee that one will develop kidney disease, there is a greater likelihood of acquiring CKD for people with diabetes. Approximately 30 percent of those with Type 1 diabetes and 10-40 percent of those with Type 2 diabetes will suffer from kidney failure. Thus, diabetes increases the likelihood that a person will suffer from CKD. High blood pressure, also known as hypertension, is another major cause of

CKD. High blood pressure refers to the high pressure of the force of blood as it pushes against the wall of blood vessels when the heart pumps blood throughout the body. High blood pressure can narrow the blood vessels within the kidneys, decreasing the flow of blood and preventing the kidneys from working properly. Similarly to diabetes, high blood pressure can prevent the kidney from removing excess fluids and wastes from the body. This excess fluid can also raise blood pressure, leading to a dangerous cycle between high blood pressure and CKD. Those who have high blood pressure and kidney disease may experience the following symptoms: loss of appetite, nausea, vomiting, drowsiness, headaches, poor concentration, increased/decreased urination, itching, numbness, weight loss, muscle cramps, chest pain, or shortness of breath (*Causes of Chronic Kidney Disease - NIDDK*, n.d.). In 2000, around a quarter of the adult population had hypertension, and this proportion is projected to increase by about 60% in 2025 (Webster et al., 2017). While diabetes and high blood pressure are the major determinants of CKD, there are many other causes that contribute to the disease. Immunoglobulin A (IGA) nephropathy can contribute to CKD. This autoimmune disease occurs when antibodies enter the kidneys and cause inflammation and kidney damage. Immunoglobulin A can injure the glomeruli which are small blood vessels within the kidney that filter blood and cause the kidneys to pass blood and protein into urine. Damage can also be done to the nephrons which are the filtering units where the glomeruli are found. Anti-glomerular basement membrane disease can contribute to CKD as it causes the immune system to create antibodies that attack the kidneys. Similar to the previous causes of CKD, anti-GBM disease can inflame the glomeruli, hindering the filtration of excess waste and fluid within the blood. However, anti-GBM disease is rare as 1 in 1 million new cases are reported a year. Alport syndrome, an inherited disorder of kidney inflammation also damages the small blood vessels in the kidneys and can lead to CKD. Additionally, hemolytic uremic syndrome occurs when red blood cells get destroyed and obstruct the kidney's filtration system, increasing the risk for CKD. Renal Artery Stenosis, another determinant of CKD, occurs when the renal arteries, the blood vessels that carry blood from the aorta to the kidneys, narrow. In addition to genetic risks, epigenetics is another determinant of CKD (Webster et al., 2017). Epigenetics refers to genetic modifications that occur without altering the DNA sequence. There is growing evidence that reveals that oxidative stress, uremia, and hyperhomocysteinemia can lead to changes in the epigenome that overlook fibrosis and may be a contributing factor in the progression of CKD. Renal fibrosis is another cause of CKD as it is characterized by the failed wound-healing of kidney tissue after an injury. Thus, the damage done to the kidney tissue increases the risk for the development of CKD. Among several other causes of CKD are heart disease, obesity, elderly age, family history of CKD, personal history of acute kidney injury, and use of tobacco products (*The Heart and Kidney Connection*, 2024).

Kidney Transplants

The first kidney transplantations in Middle Eastern countries, excluding Palestine, took place in the seventies or eighties (Rouchi & Mahdavi-Mazdeh, 2016). The International Registry in Organ Donation and Transplantation and the Global Database on Donation and

Transplantation recorded renal transplantation activity in Middle Eastern countries over 10 years beginning in 2004 and ending in 2013. Among the countries that were analyzed in this study were Pakistan, Afghanistan, Djibouti, Sudan, Yemen, Somalia, Egypt, Morocco, Syria, Iraq, Iran, Jordan, Libya, Lebanon, Turkey, Tunisia, Oman, Algeria, United Arab Emirates, Saudi Arabia, Kuwait, Qatar, and Bahrain. These countries were all divided into four groups based on their human development index (HDI). The low HDI group consisted of Pakistan, Afghanistan, Djibouti, Sudan, Yemen, and Somalia, the medium HDI group consisted of Egypt, Morocco, Syria, and Iraq, the high HDI group consisted of Iran, Jordan, Libya, Lebanon, Turkey, Tunisia, Oman and Algeria, and the very high HDI group consisted of the United Arab Emirates, Saudi Arabia, Kuwait, Qatar, and Bahrain. Prior to the study, it was proposed that the low HDI groups would face a greater burden of disease and a lack of resources compared to the middle and high HDI groups. The data recorded regarding kidney transplantations from 2004 to 2013 came from Iran, Kuwait, Lebanon, Morocco, Qatar, Saudi Arabia, Tunisia, and Turkey. Each year the kidney transplant per million population (pmp) was recorded from living donors as well as deceased donors.

Table 1 (Rouchi & Mahdavi-Mazdeh, 2016)

	Iran	Kuwait	Lebanon	Morocco	Qatar	Saudi Arabia	Tunisia	Turkey
2004	23 3	33.2 12.7	0 1	0.6 0	4 8	9.2 3.9	0 0	6.3 3.6
2005	24.5 2.9	23.2 11.4	20 0	0.3 0	5.3 2.6	9.6 4.1	0 0	7 4.1
2006	23 3.4	0 0	17.5 2	0 0	2.6 4	9.3 6.4	0 0	8.2 4
2007	22.7 4.4	0 0	0 0	0 0	1.3 5.3	10.7 4.9	0 0	10.1 5.6
2008	21.7 5.3	0 0	18.8 0	0 0	1.3 2.6	9.5 7	9.9 2.8	13.2 5.9
2009	23.5 5.4	0 0	0 0	0 0	0 2.6	11.4 4.5	0 0	27.6 6.2
2010	21.8 7.9	0 0	16 0	0 0	1.8 3.6	13.7 6.1	0 0	27.1 5.1
2011	19.7 10.1	0 0	16.2 3.8	0.4 0.1	0 0	16 7	9.2 1	33.6 7
2012	19.7 12	13.3 10	22 3	1 0.5	0 3	19 4.4	11.3 0.2	31.5 7
2013	19.5 15.1	16.7 5.3	27.3 2.5	1.2 0.9	1.5 3	17.1 3.5	10 1.7	31.5 7.8

First Row - Kidney Transplant Per Million Population from Living Donors

Second Row - Kidney Transplant Per Million Population from Deceased Donors

*0 can refer to either no activity or unreported activities

Table 1 highlights that kidney transplants per million population from 2004 to 2013 was higher from living donors than it was from deceased donors (Rouchi & Mahdavi-Mazdeh, 2016). Excluding years where both kidney transplant (pmp) from living donors and deceased donors had 0, the kidney transplant (pmp) from living donors was much greater in every country listed than it was from deceased donors.

In 2013, the kidney transplantation rates in the Middle East were 11.15 pmp compared to 31.68 in America. In 2021, 8,772 kidney transplants took place for both adult and pediatric patients in the Middle East (Saeed, 2022). Out of these transplants, 1,399 came from deceased donors. However, the deceased kidney transplant program is only available in 8 out of 18 Middle Eastern countries. Over 29 countries are part of the Middle East Society for Organ Transplantation (MESOT) (Shaheen, 2009). Among the challenges faced by members of MESOT are poor health infrastructure, inadequate government support for organ transplantation, and lack of public awareness of organ transplantation. Approximately 200 patients per million population are waiting for renal transplantation. However, the majority of patients in MESOT countries die while waiting for organs. In the data table below, kidney transplantation in the Middle East is compared with that in North America (Mudiayi et al., 2022).

Table 2 (Mudiayi et al., 2022)

	Middle East	North America
Global incidence of kidney transplantation (pmp) - Median	12.9	56.0
Global incidence of deceased donor kidney transplantation (pmp) - Median	4.4	40.6
Global incidence of living donor kidney transplantation (pmp) - Median	12.77	15.4
Kidney transplantation Waitlist - National	45%	40%
Kidney transplantation Waitlist - Regional	36%	20%
Kidney transplantation Waitlist - None	18%	40%

Table 2 highlights that the global incidence of kidney transplantation (pmp) is greater in North America (56.0) than it is in the Middle East (12.9). North America also has a greater global incidence of living and deceased donor kidney transplantation relative to the Middle East. For instance, the global incidence of deceased donor kidney transplantation is 4.4 pmp in the Middle East and 40.6 pmp in North America, and the global incidence of living donor kidney transplantation is 12.7 pmp in the Middle East and 15.4 in North America. Additionally, the kidney transplantation waitlist is greater in the Middle East both nationally and regionally, highlighting the growing number of people desiring kidney transplants in the Middle East. In

contrast, the percentage of those not on a kidney transplantation waitlist is 40% in North America while it is 18% in the Middle East.

Table 3 (Mudiayi et al., 2022)

	Renal Allograft Function (%)					Graft Survival (%)					Patient Survival (%)				
	0	1-10	11-50	51-75	>75	0	1-10	11-50	51-75	>75	0	1-10	11-50	51-75	>75
Middle East	0	0	0	27	73	0	0	9	27	64	0	0	9	9	82
N. America	0	20	0	0	80	0	20	0	0	80	0	0	20	0	80

Definitions:

Renal Allograft Function - Health of a transplanted kidney in a patient

Graft Survival - Amount of time a transplanted kidney is functional

Patient Survival - Percentage of patients who survive following a kidney transplant

The percentage of patients who have a renal allograft function above 75% and graft survival above 75% is greater in North America than it is in the Middle East. However, the percentage of patients who have a survival rate above 75% is slightly greater in the Middle East than it is in North America. Despite having a slightly lower patient survival rate than in the Middle East, patient outcome is much better in North America as highlighted by the renal allograft function and graft survival which is greater than in the Middle East.

Table 4 (Mudiayi et al., 2022)

	GDP (PPP), \$ billion	Total health expenditures (% of GDP)	Annual Cost KRT (\$) and out-of-pocket/% paid by patient from total cost KT (first year)
Global - median	130,483,015	6.2	26, 903
Middle East - median	5,478,576	4.5	19, 258

*GDP - Gross Domestic Product based on purchasing power parity

When compared to the global median, the Middle East median is lower in regard to total health expenditures. The Middle East spends a lower percentage of its GDP on total health expenditure, sitting at 4.5% compared to the global median of 6.2%. The United States on the other hand spends 17.3% of its GDP on health expenditures with its GDP sitting at roughly \$27.36 trillion as of 2023 (*World Bank Open Data*, n.d.). Moreover, the cost of a kidney

transplant in year 1 is about \$120,203 (Jha et al., 2023). However, insurance with Medicaid covers 60% while private insurance covers 32% (Wang & Hart, 2021). Thus, the United States costs much more for kidney transplants when compared to the Middle East despite spending a greater percentage of its GDP on health expenditures.

Dialysis Modalities

As previously mentioned, hemodialysis is a demanding process as a patient is required to come to a dialysis unit around 3 times a week with each session lasting approximately 4 hours (Zazzeroni et al., 2017). Those under hemodialysis have fluid and food restrictions and are required to take various medications. On the other hand, peritoneal dialysis can take place at a patient’s home multiple times a day.

Table 5 (Mudiayi et al., 2022)

	Official Registry	
	Dialysis (%)	Transplantation (%)
Middle East	55	73
North America	44	33

While dialysis modalities do not take up the majority of kidney treatment in the Middle East, they still play a significant role as it takes up 55%. On the other hand, in North America, dialysis modalities have a greater prevalence than kidney transplantation.

Table 6 (Karam et al., 2024)

	Prevalence of long-term dialysis, pmp		Long-term dialysis centers, pmp	
	HD	PD	HD	PD
Global, median	322.7	21.0	5.1	1.6
Middle East, median	276.6	17.5	3.3	0.6

Table 6 reveals that the prevalence of long-term dialysis pmp for both hemodialysis and peritoneal dialysis is less in the Middle East than it is globally. Hemodialysis is evidently more used worldwide as the prevalence of long-term dialysis pmp for the global median is 322.7 while that for the Middle East median is 276.6. On the other hand, the prevalence of long-term dialysis pmp for the global median is 21.0 while that for the Middle East median is 17.5. These trends are also reflected in the long-term dialysis centers pmp as the Middle East has lower hemodialysis and peritoneal dialysis centers. The United States currently has approximately 7,578 dialysis centers treating around 519,241 transplant centers (*National ESRD Census Data*, n.d.) with a greater pmp for both hemodialysis peritoneal dialysis centers than that in the Middle East.

Table 7 (Karam et al., 2024)

	Annual cost KRT (\$) and out-of-pocket/% paid by patient from total cost	
	HD	PD
Global - median	19,380	18,959
Middle East - median	26,226	16,136

Table 7 highlights that patients in the Middle East pay more for hemodialysis compared to the global median but less for peritoneal dialysis. Similar to kidney replacement in the United States, the cost of hemodialysis and peritoneal dialysis depends on insurance. As of 2024, the average annual cost for hemodialysis for 15 required sessions without Medicare is \$88,000 but \$17,600 with Medicare (*Dialysis Cost With and Without Insurance in 2024*, n.d.). The average cost for peritoneal dialysis for 25 required sessions is \$71,000 without Medicare but \$14,000 with Medicare. Nevertheless, the cost for dialysis modalities in the United States is much greater than in the Middle East but insurance coverage creates a variation in the percentage paid by patients from the total cost.

ANALYSIS

Based on the data from 2004 to 2013 in the countries Iran, Kuwait, Lebanon, Morocco, Qatar, Saudi Arabia, Tunisia, and Turkey, the Middle East relies more heavily on living donors than deceased donors (Rouchi & Mahdavi-Mazdeh, 2016). While those who acquire kidneys from living donors do indeed have better graft survival, meaning that the kidney will function for a longer period of time, the utilization of kidneys from deceased donors will increase the availability of kidneys overall. By having greater access to kidneys from both living and deceased donors, a greater number of patients will be able to receive kidney transplants. Some may argue that living donors are preferred to deceased donors due to the longevity and efficiency of the kidneys of living donors. However, patients with CKD are at risk of death due to a failing kidney which does not perform its regular function of removing metabolic waste products and filtering blood. Thus, the utilization of kidneys from both living and deceased donors will increase the availability of kidney organs and allow for more patients to receive functioning kidneys.

When compared to North America, the Middle East has a significantly lower global incidence of kidney transplantation (pmp). The global incidence of kidney transplantation refers to the number of kidney transplants performed per million population per year. North America having a global incidence of kidney transplantation (pmp) of 56.0 and the Middle East having one of 12.9 highlights that North America performs a significantly higher number of kidney transplants per year. Based on the data in Table 2, North America utilizes a greater number of kidneys from both living and deceased donors when compared to the Middle East. The range between the median number of living donor kidney transplants between North America (15.4) and the Middle East (2.7) is not significant. However, the range between the median number of

deceased donor kidney transplants between North America (40.6) and the Middle East (4.4) is significant. This great range demonstrates that North America performs more kidney transplants a year partly because it utilizes kidneys from both living and deceased donors. In the Middle East, the majority of kidney transplants come from living donors. By heavily focusing on kidneys from living donors rather than on kidneys from both living and deceased donors, the Middle East limits the amount of people who can acquire a kidney transplant. If the Middle East increases the number of kidneys they use from deceased donors, the global incidence of kidney transplantation in that region will increase.

The waitlist for kidney transplantation is also greater in the Middle East than in North America with the percentage of those not on the waitlist for kidney transplantation being far less in North America. The greater percentage of those on the kidney waitlist in the Middle East highlights that the heavy reliance on kidneys from living donors plays a role. Because the Middle East prioritizes kidneys from living donors, it not only limits its global incidence of kidney transplantation but also increases the percentage of patients on the kidney transplantation waitlist. This hinders its potential for the amount of kidney transplants that can be performed a year.

In Table 3, the patient outcomes for those who received kidney transplants in North America and the Middle East are revealed. North America has a greater percentage of patients who have a renal allograft function of 75% or higher and a greater percentage of patients who have a graft survival of 75% or higher. A higher renal allograft function and the health of a transplanted kidney in a patient provide insight into the more efficient treatment that patients in North America receive than those in the Middle East. Despite utilizing kidneys from both living and deceased donors, the renal allograft function of patients in North America is still much better than those in the Middle East. Kidneys from living donors are associated with greater renal allograft function compared to kidneys from deceased donors. Although the Middle East utilizes more kidneys from living donors than it does from deceased donors, it has a lower renal allograft function than North America, demonstrating the more efficient kidney treatment in North America.

In addition, the percentage of patients who have a graft survival of 75% or higher in North America is greater than in the Middle East, again demonstrating the more efficient kidney treatment in North America. The graft survival, the amount of time a transplanted kidney is functional, reveals that the Middle East does not provide equivalent kidney treatment to its patients as North America does. Therefore, the patient outcomes in North America are much better than in the Middle East as North America has a greater percentage of patients who have a renal allograft function and a graft function of 75% or greater. However, the Middle East has a slightly higher percentage of patients who have a survival rate of 75% or greater. For instance, 82% of kidney transplant patients in the Middle East have a survival rate of 75% or greater while 80% of those patients in North America have the same survival rate. Nevertheless, this does not highlight that North America is less superior to the Middle East in kidney transplants.

North America performs many more kidney transplants than the Middle East as seen by its global incidence of kidney transplantation, so it is bound that North America has a slightly lower percentage of patients with a survival rate of 75% or greater.

Furthermore, North America, particularly the United States, spends a greater percentage of its GDP on health expenditures compared to the Middle East. While the GDP spent on health expenditures may be used for advanced health technologies and nephrologists, kidney transplants remain much higher in the United States. Although Medicaid and other health insurance companies may decrease the cost that patients pay out of pocket, the cost of kidney transplantation remains much higher in the United States. Therefore, the cost of kidney transplants is lower in the Middle East. The affordability of kidney transplants in the Middle East depends on the income of patients and the parts of the Middle East that the patients live in, but overall, kidney transplants cost much less in the Middle East than in the United States.

In both the Middle East and North America, the prevalence of hemodialysis is much greater than that of peritoneal dialysis. The same trend is also found in dialysis centers as both the Middle East and North America have more hemodialysis centers than peritoneal dialysis centers. However, similar to kidney transplants, North America utilizes much more dialysis modalities, such as hemodialysis and peritoneal dialysis, than the Middle East. This reaffirms that kidney treatment in North America is more efficient than kidney treatment in the Middle East.

The Middle East can increase the usage of dialysis modalities to enhance its kidney treatment. Hemodialysis requires in-person centers while peritoneal dialysis can be performed at a patient's home. By increasing the number of hemodialysis centers and increasing the accessibility of peritoneal dialysis, a greater number of patients will be able to acquire kidney treatment without requiring a kidney transplant. Similar to kidney transplants, the cost of dialysis for both hemodialysis and peritoneal dialysis is greater in North America, particularly in the United States. Again, different health insurance can reduce the overall cost of these dialysis modalities. Regardless, the cost of hemodialysis and peritoneal dialysis is much lower in the Middle East than in North America.

Future Directions

It is imperative that the countries in the Middle East take immediate action to improve the conditions of those suffering from CKD. The results of the present study suggest that the Middle East can improve kidney treatment for its patients by utilizing kidneys from deceased donors in kidney transplants. Although many may argue that kidneys from live donors last longer and are better for patients, they fail to recognize the inefficiency that arises from heavily relying on these live donors. Because the Middle East relies heavily on kidneys from live donors, its global incidence of kidney transplants is much lower than that of North America. Many studies suggest that increasing the number of kidney transplants from deceased donors will allow a greater number of people to receive kidney treatment. Consequently, the global incidence of kidney transplants in the region would increase while the waiting list for these transplants would

decrease. Additionally, these studies highlight that the Middle East could allocate a greater percentage of its GDP to health expenditures. Particularly, these funds could establish more dialysis centers to expand the accessibility of dialysis modalities, such as hemodialysis or peritoneal dialysis. However, challenges can arise in spending a greater percentage of its GDP on these health facilities as Middle Eastern countries may prioritize using its funds on other economic or political aspects. Challenges may also arise in regard to the affordability of these forms of kidney treatment for impoverished families.

Conclusion

Chronic Kidney Disease is a life-threatening disease that can impair the daily function of those involved. CKD consists of five different stages with the final stage, end-stage kidney disease, posing the greatest threat to one's health. Among some of the determinants of CKD are diabetes, high blood pressure (hypertension), Immunoglobulin A nephropathy, Anti-glomerular basement membrane, Alport syndrome, hemolytic uremic syndrome, renal artery stenosis, oxidative stress, uremia, hyperhomocysteinemia, renal fibrosis, heart disease, obesity elderly age, family history of CKD, personal history of acute kidney injury, use of tobacco products, etc. In regards to kidney treatment, the accessibility and utilization of kidney transplantation and dialysis modalities greatly impact patient outcomes in the Middle East. Middle Eastern countries utilize more kidneys from live donors than deceased ones. Compared to North America, the Middle East has a lower global incidence of kidney transplantation (pmp), and a global incidence of deceased and living donor kidney transplantation but a greater kidney transplantation waitlist. Additionally, the percentage of patients who have renal allograft function and graft survival above 75% is greater in North America than in the Middle East. On the other hand, the Middle East has a slightly higher patient survival than North America which can be accounted for by the lower number of kidney transplants performed by regions in the Middle East. Furthermore, the cost of kidney transplants is greater in North America but varies due to medical insurance. However, the Middle East spends a lower percentage of its GDP on total health expenditures. Furthermore, when compared to the United States, the Middle East has a lower prevalence of long-term dialysis pmp and dialysis centers for both hemodialysis and peritoneal dialysis. Hemodialysis requires a patient to go to a dialysis center two or three times with each session lasting around three to four hours while peritoneal dialysis can take place at a patient's home multiple times a day. Similar to kidney transplants, the Middle East has a much lower cost for both hemodialysis and peritoneal dialysis relative to the United States. The Middle East can greatly improve its kidney treatment by utilizing more kidneys from deceased donors, spending more of its GDP on health expenditures, increasing the prevalence of hemodialysis and peritoneal dialysis, and expanding the number of centers accessible.

Works Cited

- Ammirati, A. L. (2020). Chronic Kidney Disease. *Revista Da Associação Médica Brasileira*, 66, s03–s09. <https://doi.org/10.1590/1806-9282.66.S1.3>
- Andreoli, M. C. C., & Totoli, C. (2020). Peritoneal Dialysis. *Revista Da Associação Médica Brasileira*, 66, s37–s44. <https://doi.org/10.1590/1806-9282.66.S1.37>
- Augustine, J. (2018). Kidney transplant: New opportunities and challenges. *Cleveland Clinic Journal of Medicine*, 85(2), 138–144. <https://doi.org/10.3949/ccjm.85gr.18001>
- Causes of Chronic Kidney Disease—NIDDK. (n.d.). National Institute of Diabetes and Digestive and Kidney Diseases. Retrieved July 29, 2024, from <https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/causes>
- Diabetes—A Major Risk Factor for Kidney Disease. (2015, December 24). National Kidney Foundation. <https://www.kidney.org/atoz/content/diabetes>
- Dialysis Cost With and Without Insurance in 2024. (n.d.). Mira Health. Retrieved August 7, 2024, from <https://www.talktomira.com/post/how-much-does-dialysis-cost-in-2022>
- Ferguson, M. A., & Waikar, S. S. (2012). Established and Emerging Markers of Kidney Function. *Clinical Chemistry*, 58(4), 680–689. <https://doi.org/10.1373/clinchem.2011.167494>
- Himmelfarb, J., & Ikizler, T. (2010). Hemodialysis. *The New England Journal of Medicine*, 363, 1833–1845. <https://doi.org/10.1056/NEJMra0902710>
- Jha, V., Al-Ghamdi, S. M. G., Li, G., Wu, M.-S., Stafylas, P., Retat, L., Card-Gowers, J., Barone, S., Cabrera, C., & Garcia Sanchez, J. J. (2023). Global Economic Burden Associated with Chronic Kidney Disease: A Pragmatic Review of Medical Costs for the Inside CKD Research Programme. *Advances in Therapy*, 40(10), 4405–4420. <https://doi.org/10.1007/s12325-023-02608-9>
- Karam, S., Amouzegar, A., Alshamsi, I. R., Al Ghamdi, S. M. G., Anwar, S., Ghnaimat, M., Saeed, B., Arruebo, S., Bello, A. K., Caskey, F. J., Damster, S., Donner, J.-A., Jha, V., Johnson, D. W., Levin, A., Malik, C., Nangaku, M., Okpechi, I. G., Tonelli, M., ... Savaj, S. (2024). Capacity for the management of kidney failure in the International Society of Nephrology Middle East region: Report from the 2023 ISN Global Kidney Health Atlas (ISN-GKHA). *Kidney International Supplements*, 13(1), 57–70. <https://doi.org/10.1016/j.kisu.2024.01.009>
- Mudiayi, D., Shojai, S., Okpechi, I., Christie, E. A., Wen, K., Kamaleldin, M., Elsadig Osman, M., Lunney, M., Prasad, B., Osman, M. A., Ye, F., Khan, M., Htay, H., Caskey, F., Jindal, K. K., Klarenback, S., Jha, V., Rondeau, E., Turan Kazancioglu, R., ... Bello, A. K. (2022). Global Estimates of Capacity for Kidney Transplantation in World Countries and Regions. *Transplantation*, 106(6), 1113–1122. <https://doi.org/10.1097/TP.0000000000003943>
- National ESRD Census Data. (n.d.). Retrieved August 7, 2024, from <https://esrdnetworks.org/resources-news/national-esrd-census-data/>
- Rouchi, A. H., & Mahdavi-Mazdeh, M. (2016). Kidney transplantation in the Middle East. *Clinical Nephrology*, 86 (2016)(13), 101–105. <https://doi.org/10.5414/CNP86S109>
- Saeed, B. (2022). Pediatric Kidney Transplantation in the Middle East: Challenges and Solutions. *Experimental and Clinical Transplantation: Official Journal of the Middle East*

- Society for Organ Transplantation, 20(Suppl 3), 7–14.
<https://doi.org/10.6002/ect.PediatricSymp2022.L2>
- Shaheen, F. A. M. (2009). Organ donation in the Middle East countries. *Ethnicity & Disease*, 19(1 Suppl 1), S1-16–17.
- The Heart and Kidney Connection. (2024, April 4). National Kidney Foundation.
<https://www.kidney.org/atoz/content/heart-and-kidney-connection>
- Wang, J. H., & Hart, A. (2021). Global Perspective on Kidney Transplantation: United States. *Kidney360*, 2(11), 1836–1839. <https://doi.org/10.34067/KID.0002472021>
- Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. *The Lancet*, 389(10075), 1238–1252. [https://doi.org/10.1016/S0140-6736\(16\)32064-5](https://doi.org/10.1016/S0140-6736(16)32064-5)
- World Bank Open Data. (n.d.). World Bank Open Data. Retrieved August 7, 2024, from <https://data.worldbank.org>
- Zazzeroni, L., Pasquinelli, G., Nanni, E., Cremonini, V., & Rubbi, I. (2017). Comparison of Quality of Life in Patients Undergoing Hemodialysis and Peritoneal Dialysis: A Systematic Review and Meta-Analysis. *Kidney and Blood Pressure Research*, 42(4), 717–727. <https://doi.org/10.1159/000484115>

**Investigating Carbon Markets:
Mechanisms, Economics, and Implementation Strategies for a Sustainable Future
By Timothy Liang**

Abstract

In the relentless battle against climate change, the 1997 Kyoto Protocol emerged as a beacon of hope, forging a pact between nations to limit and reduce greenhouse gas emissions (United Nations, n.d.). A decade later, the Paris Agreement was created, aiming to combat climate change through global cooperation. Its main goal is to limit the global temperature increase below 2°C, as well as provide financing for developing countries' climate action. It was here that carbon markets emerged as crucial mechanisms for reducing greenhouse gas emissions. This project explores the function of various carbon markets across different countries, examining their regulatory frameworks, trading mechanisms, and impact on emission reduction. Through a deep dive into carbon market economics, this analysis examines how these markets incentivize emission reductions through the trading of carbon credits. Carbon credits are units that represent the removal of one metric ton of carbon from the atmosphere. This paper assesses carbon credit valuation, considering factors such as market demand, regulatory policies, and project quality. Furthermore, it explores challenges in establishing a global carbon market, including regulation, structure, and international politics. By analyzing existing models and identifying best practices, the project aims to propose an ideal carbon market model that fosters international cooperation and emission reduction at scale. The proposed model ensures long-term climate resilience through reliable and impactful carbon market strategies.

Introduction

Climate change is an undeniable reality, impacting every corner of the globe. Rising temperatures lead to more frequent and intense heat waves, causing widespread crop failures and significant shifts in ecosystems. Consequently, crop supplies are becoming increasingly unstable, leading to food shortages and price spikes. For example, premature budding due to a warm winter led to \$220 million in damages for 2012 Michigan cherry crops (EPA, n.d.). Additionally, if current emission trends continue, the hottest daily temperatures could increase by at least 10°F in the United States alone by the end of the century (Union of Concerned Scientists, n.d.).

In addition, this warming climate is causing glaciers to melt, contributing to rising sea levels. Coastal communities are increasingly at risk as the global mean sea level rises at least 0.3 meters (1 foot) above 2000 levels by 2100 (Lindsey, 2022). These rising sea levels pose existential threats to coastal ecosystems, leading to increased flooding and habitat loss.

Extreme weather events are becoming more severe and frequent. Mega-storms like Hurricane Harvey, previously considered once-in-a-lifetime century events, now strike approximately every 16 years (Union of Concerned Scientists, n.d.). This increase in extreme weather events causes widespread damage, leading to loss of life and also increased financial burden.

Climate change's increasing severity is directly linked to the rise in carbon emissions. In 2023, global carbon emissions surged to record levels, surpassing 40 billion tons (Stanford Doerr, 2023). Current emission levels suggest that we will exceed the Paris Agreement's 1.5°C temperature limit by 2030 (Stanford Doerr, 2023). As our carbon budget depletes, the urgency to achieve net zero emissions intensifies. This necessitates intensified carbon dioxide removal efforts even after reaching global net-zero. This escalation underscores the importance of effective carbon market strategies and highlights their critical role in shaping our environmental future.

International efforts have played a crucial role on the road to reduce carbon emissions. For example, the Kyoto Protocol, aimed primarily at industrialized countries, established emission reduction targets for 37 countries and the European Union in 1997 (United Nations, n.d.). Adhering to the principle of "common but differentiated responsibility," it also provided assistance for the Adaptation Fund, which helped developing countries react to climate change (United Nations, n.d.).

Building on this foundation, the Paris Agreement emerged as a landmark for global cooperation in combating climate change. Aiming to limit the global temperature rise to below 1.5°C, the agreement emphasizes reducing greenhouse gas emissions and provides financing for climate action in developing countries. By requiring countries to submit updated emission reduction plans every five years, it provides a framework for ongoing global climate action and regular progress assessment (United Nations, n.d.).

Leading these global efforts is the Conference of the Parties (COP), the highest decision-making body under the United Nations Framework Convention on Climate Change (UNFCCC). COP ensures effective institutional and administrative arrangements to combat climate change by reviewing the implementation of climate agreements (University of Cambridge, n.d.). During its annual meetings, the COP reviews national communications and emission inventories submitted by parties, guiding the assessment of measures and progress.

The Intergovernmental Panel on Climate Change (IPCC) supports these initiatives, with top climate scientists dedicated to delivering comprehensive assessments of climate science. Through its authoritative reports, the IPCC has established a scientific consensus on the link between human activities and climate change, guiding the implementation of crucial agreements like the Paris Agreement (IPCC, n.d.).

Complementing these international efforts, carbon markets have become essential tools in the global fight against climate change. These markets facilitate the trading of carbon credits, allowing countries and companies to offset their emissions and support reduction efforts.

Carbon markets are broadly classified into two types: compliance and voluntary. Compliance carbon markets operate under government or regional regulations, requiring companies to meet emission limits or face penalties. These markets, often using cap-and-trade mechanisms, are crucial for meeting emission reduction targets set by the Kyoto Protocol and Paris Agreement (Carbon Credits, n.d.). Voluntary carbon markets (VCMs), in contrast, depend

on voluntary participation, enabling participants to choose to offset their emissions by purchasing carbon credits.

Carbon credits are also categorized into two main types: avoidance and removal. Avoidance credits are generated by projects that prevent the release of greenhouse gasses, such as renewable energy projects and REDD+ forestry and farming initiatives (Favasuli & Sebastian, 2021). These credits are priced based on factors like project location, vintage, and their potential to provide additional social and environmental benefits. On the other hand, removal credits are generated by projects that actively capture and sequester greenhouse gasses. Examples include nature-based projects like reforestation and technological solutions like direct air capture and carbon capture and storage (CCS) (Favasuli & Sebastian, 2021). Removal credits often command higher prices due to the significant investment required and their perceived effectiveness in combating climate change.

The key difference is that avoidance credits prevent emissions, while removal credits actively remove emissions from the atmosphere. Avoidance credits typically trade at lower prices, from a few cents to a few dollars per metric ton of CO₂ equivalent (mtCO₂e). Nature-based removal projects may trade between \$15/mtCO₂e to \$20/mtCO₂e, while technological removal projects like CCS can reach prices of \$100/mtCO₂e or more (Favasuli & Sebastian, 2021). Prices for both types of credits are negotiated between buyers and sellers in over-the-counter deals or through exchanges.

Common problems associated with carbon finance include issues of honesty, where projects may not genuinely result in emissions reductions if they are already legally required or economically viable without carbon finance. Leakage presents another challenge, as emissions reductions achieved in one area may be offset by increased emissions in another area not covered by the project, undermining the overall effectiveness. Accuracy is crucial as different types of projects require different methodologies and using inappropriate methods can lead to overestimation or underestimation of reductions. Furthermore, the risk of double counting emissions reductions poses a significant concern, as the same reductions may be claimed by multiple parties, leading to inaccuracies in overall emissions.

In summary, addressing climate change requires a multifarious approach involving international cooperation, strict policies, and effective market mechanisms.

Carbon Market Current Status

After a period of strong initial growth, the market faced stagnation and decline following the global economic and financial crisis of 2008/09 (Kreibich & Hermwille, 2021). This downward trajectory continued post-Paris Agreement, reflecting uncertainties around the viability and legitimacy of voluntary offsetting.

In recent years, things have changed. In 2018 and 2019, the voluntary carbon market rebounded in both market volume and value, shifting from public institutions like the World Bank to private companies. This shift has been driven by numerous large corporations making ambitious net zero commitments, signaling a new era for the market. An analysis of 482 large

companies reveals the magnitude of this shift, with collective annual revenues surpassing \$16 trillion for global VCM's (Kreibich & Hermwille, 2021). These companies are crucial in shaping the future of the voluntary carbon market through their carbon offset commitments. While a minority explicitly exclude offsets, many plan to offset their remaining emissions. Carbon markets are complex and vary widely. To create a more reliable and effective future carbon market, understanding the main models is essential.

Types of Carbon Markets

Carbon markets are primarily divided into two models: compliance markets and voluntary carbon markets. Each model has its own set of benefits and challenges, playing crucial roles in tackling climate change.

Compliance Markets

Compliance carbon markets set mandatory limits on carbon emissions for companies, typically using a cap-and-trade system. These markets ensure that emission reduction targets are always met through Nationally Determined Contributions (NDCs), which were set out in the Paris Agreement in 2015 (Archer, 2023). These contributions are periodically reevaluated and updated by the participating countries.

Cap and trade systems offer economic incentives for reducing carbon emissions. The government sets a cap on the total amount of a specific pollutant that can be emitted. Companies or other organizations are issued emission permits and are required to hold an equivalent number of credits to cover their emissions. The total number of credits cannot exceed the cap set, which limits total emissions to the cap.

Companies that need to increase their emission allowance must buy more credits from those who emit less, thus creating a financial incentive for companies to reduce emissions (Goulder & Schein, 2013). These policies have driven significant emissions reductions. For instance, the EU cap and trade system reduced emissions by 3% in Phase I (2005-2007) and by 25-28% in covered firms during early Phase II (2008-2012) (Eden et al., 2018). The EU ETS, held a few thousand companies to these caps, among those being power plants, aviation companies, and more in the industrial sectors. The reductions noted during Phase I (3%) and early Phase II (25-28% for covered firms) were substantial for the sectors directly involved. However, the impact on total national emissions varied by country depending on the share of covered sectors within the national economy.

Cap-and-trade systems also improve market stability, leading to more stable carbon prices (International Emissions Trading Association). The fixed cap on total emissions ensures a limited supply of allowances, helping maintain consistent carbon prices and create a predictable market. This predictability allows companies to anticipate the rising cost of emissions. This stable cost reduces compliance cost and creates investor confidence which drives innovation in emission reduction strategies.

Cap-and-trade systems may not effectively drive innovation in clean energy technologies due to their propensity for loopholes. One of the most common loopholes in cap-and-trade systems is the ability for companies to store emission permits for future use. A prominent example of this is in California's cap-and-trade program, where companies had stored over 200 million permits by the end of 2018 (Song, 2019). This amount was almost equivalent to the total reductions actually expected from the program from 2021 to 2030. They bank permits to save them for periods where emission caps become stricter, thereby avoiding the intended emission reductions when caps were looser (Song, 2019).

Cap-and-trade systems often face political resistance thus, policymakers face significant pressure from industries who are affected by increased costs. This resistance can lead to watered-down policies that fail to set the correct caps needed for meaningful emissions reductions (Nordhaus & Shellenberger, 2009).

In summary, cap-and-trade systems are complex, making them difficult to manage effectively. This complexity leads to loopholes and opportunities for corruption. Companies exploit these loopholes to continue emitting greenhouse gasses without significant penalties. As a result, the system may fall short in promoting substantial technological advancements needed for long-term climate solutions (Nordhaus & Shellenberger, 2009).

Voluntary Markets

In contrast, voluntary carbon markets (VCM) enable companies and organizations to purchase carbon credits to offset their emissions. These markets are driven by consumer demand and corporate responsibility rather than government mandates (UNDP, n.d.). These markets are entirely voluntary, with many companies participating to demonstrate social responsibility or enhance their brand image. Companies that purchase carbon credits can enhance their corporate image and customer loyalty, attracting environmentally conscious consumers and boosting sales, leading to further growth of the voluntary carbon market (Carbon Credits, n.d.).

The growth of the voluntary carbon market is based on supply and demand. As more companies commit to reducing their carbon footprint, the demand for carbon credits increases, driving up their price and increasing revenue potential. This not only directly generates more revenue for carbon offset projects, but it also inspires new innovative and cost-effective methods for generating credits, enhancing profitability through technological advancements (Carbon Credits, n.d.).

Voluntary markets are also flexible, allowing for participants to make their own choices. Voluntary markets allow companies to demonstrate their commitment to sustainability and enhance their corporate image as they volunteer to offset emissions. However, this flexibility leads to notable drawbacks in VCMs. In January 2023, a report by *TheGuardian*, *Die Zeit*, and *SourceMaterial* assessed credits certified by Verra, a leading carbon standard. The investigation revealed that over 90 percent of their forestry credits did not represent real emissions reductions (Dawes, 2024). This led to carbon credit prices falling sharply due to growing doubts about the effectiveness and integrity of nature-based projects.

VCMs may also allow companies to continue polluting by using carbon credits to falsely claim emission reductions. Many carbon credits sold in VCMs lack proper verification. This means that the projects they are supposed to support may not be genuinely reducing emissions. For example, a company might buy credits from a reforestation project that claims to plant trees, but without proper oversight, it's difficult to ensure that the claims are true. These credits are often a cheap fix, letting companies label themselves as "carbon neutral" without any real verification. Since ownership and details of many credits remains mysterious, it is easy for companies to make unverifiable claims (Global Witness, 2022).

Another major problem with the voluntary carbon market is the flawed calculation of carbon credits. These calculations often rely on complex and sometimes inaccurate methodologies to estimate the amount of carbon emissions reduced or sequestered by various projects. For instance, a reforestation project might claim to sequester a certain amount of CO₂ based on tree growth predictions, but these estimates can be overly optimistic or fail to account for unforeseen events like forest fires or disease.

Furthermore, most credits originate from unreliable forestry and non-additional renewable energy projects (Global Witness, 2022). The VCM's lack of standardization and transparency further complicates things, as even reputable agencies face issues.

Carbon Tax Model

The carbon tax model, which involves a direct tax on carbon emissions, achieves economic efficiency and significant emissions reductions by providing a clear price signal for businesses. When a fixed cost is associated with emitting carbon, companies are incentivized to invest in cleaner technologies and energy efficiency to reduce their tax burden, leading to more sustainable long-term business investments. This cost framework enables businesses to plan and allocate resources effectively towards low-carbon solutions. The revenue generated provides funding so that governments can reinvest into further driving emissions reductions.

In addition, the simplicity of a carbon tax makes it much easier to implement and manage compared to a traditional cap-and-trade system (Metcalf, 2008). It is easier to manage since a carbon tax is more stable and predictable pricing compared to cap-and-trade, which reduces volatility and uncertainty. Carbon taxes are simple, but they can also be flexible. Carbon taxes can be adjusted in price based on industry. For industries that compete globally, such as agriculture and forestry, carbon taxes would be lower in order for countries to be able to compete globally to not collapse the domestic economy.

One example is Sweden's carbon tax: since being introduced in 1991, it has been successful in reducing emissions. Initially, the tax was \$44.37 per metric ton of CO₂. By 1993, industries like agriculture and forestry paid \$11.28 per metric ton, while other consumers paid \$45.15 per metric ton (Sumner, J., Bird, L., & Dobos, H. 2009). The tax rate is determined by factors such as the need to incentivize emission reductions, the economic impact on different industries, and the government's environmental and revenue goals. They also adjust the rate over time to balance environmental effectiveness and economic competitiveness.

Such industries in countries with lower carbon taxes can produce goods more cheaply, making them more appealing than goods from countries with higher environmental costs. This leads to market gain for domestic industries since they have the advantage in price competition. This pricing advantage made domestic goods cheaper, discouraging investment in countries with higher regulatory costs and leading to economic decline (Omolere, 2024).

By 2003, the standard tax rate in Sweden had increased to \$104.83 per metric ton (Sumner, J., Bird, L., & Dobos, H. 2009). The revenue from the carbon tax generated approximately \$3.65 billion annually in the mid-2000s. The tax effectively reduced CO₂ emissions, with national greenhouse gas emissions decreasing by over 40% since the mid-1970s (Sumner, J., Bird, L., & Dobos, H. 2009).

However, imposing a carbon tax faces many challenges. Imposing a carbon tax can disincentivize the creation of even more innovative technologies. The U.S. has seen little progress in passing a carbon tax. Many conservatives oppose carbon taxes due to potential mismanagement of revenues and negative impacts on fossil fuel workers and consumers. In contrast, liberals worry about disproportionate harm to low-income individuals from higher prices (Ho, 2021).

Economists Kenneth Gillingham and James Stock find we already have dozens of existing policies that place high implicit prices on carbon reductions: e.g. renewable portfolio standards that regulate electricity (with an implicit carbon price of \$0-\$190/ton), tax credits for solar power (\$140-\$2100/ton) or wind (\$2-\$260/ton), fuel economy standards (\$48-\$310/ton), corn ethanol standards (\$-18 to \$310/ton), or subsidies for electric cars (\$350-\$640/ton). The additional impact of a carbon tax at, say, the \$51/ton social cost of carbon recently adopted by the Biden administration could have a smaller effect on these specific sectors relative to the policies already in place.

There are also various examples of less successful taxes. In 2018, France's carbon tax led to violent "yellow vest" protests against rising prices, causing the French government to back down (Ho, 2021). In fact, no U.S. state has successfully passed a carbon tax (although many U.S. states and the E.U. have passed cap and trade policies). Texas Senator Cornyn states that "Over the last couple of years, we've seen no shortage of unrealistic and downright harmful policies that are advocated for in the name of reducing carbon emissions. The solution to a rising climate isn't job-killing government regulations – it's innovation and American ingenuity." Backed by the mass of U.S. senators, Cornyn argued the common idea that imposing a carbon tax will do more harm than good for the U.S (Cassidy, 2023).

A Modern Carbon Market

China, as the largest developing country and carbon emitter, has made significant efforts to establish and expand its carbon market, aiming for a national unified carbon emissions trading system. The country is working towards a national unified carbon emissions trading market to create a solid foundation for managing carbon emissions, incentivizing reductions, and integrating carbon finance into the broader economic system.

China's carbon market has three key components: trading systems, policy frameworks, and institutional arrangements (Zhou & Li, 2019). China's trading systems act as smaller components of the national market, facilitating the trading of carbon credits for participants. The government has set emission reduction targets and established regulatory frameworks to promote carbon trading. The Ministry of Ecology and Environment (MEE) oversees the carbon market to ensure compliance with current regulations (Lam & Wong et al., 2023).

Although only formed in 2021, China's carbon market has begun raising awareness among companies about carbon emissions, encouraging better monitoring, management, and reporting, with potential for significant future impact. The market has raised some awareness among companies about carbon emissions, encouraging them to accurately and effectively monitor, manage, and report their emissions (Xiaoying, 2023). Since the market's inception, there have been millions of tonnes of CO₂ traded showing active participation from companies with a high compliance rate (Zhou & Li, 2019).

Despite progress, China's carbon market faces several challenges. There is still a general lack of awareness and understanding of carbon finance, limiting market growth and participation. Additionally, underdeveloped legal frameworks limit enforcement, leading to potential compliance issues.

In addition, since there is no national emission cap, there may be weak demand for emission allowances or offsets. The majority of domestic carbon trades are driven by voluntary commitments rather than mandatory caps, which creates a stronger potential for market volatility and drops in demand (Lo, 2015).

To address these challenges, new policies for the future development of carbon finance in China have been proposed. Strengthening regulatory systems by providing clear guidelines is key to market stability. Additionally, initiatives to raise awareness and educate market participants about the mechanism and benefits of carbon finance will help boost market participation and growth.

In contrast, China's efforts to develop carbon markets are constrained by structural economic and political barriers. For example, China's carbon markets are developing in a context very different from the liberal-capitalist regimes where such markets are typically successful. The state's control over critical industries and centralized political power contrasts with the bottom-up force that drives carbon markets in liberal democracies. This centralization could undermine the decentralized economic and market institutions needed for effective carbon markets (Lo, 2015). However, this structure also presents opportunities to design a system tailored to China's unique context.

In summary, consistent with previous challenges, a functioning carbon market needs rigorous monitoring, reporting, and verification mechanisms, which are currently underdeveloped/ The absence of a consistent and accurate system undermines the effectiveness of the carbon market.

Carbon Credit Valuation

Carbon credit valuation assesses the worth of credits representing a ton of carbon dioxide or equivalent greenhouse gas reduced, sequestered, or avoided. This valuation process is critical in carbon offset markets and often involves several key components. First, fundamental to carbon credit valuation is establishing a baseline scenario. This represents the expected emissions without a carbon offset project. The difference between baseline and actual emissions after the project equals the carbon credits generated (Ratnatunga & Jones et al., 2011). Accurate measurement and monitoring of GHG reductions are essential for proper valuation. Verification methods such as field surveys or remote sensing are typically used to ensure credible and quantifiable reduction reports (Badgley & Freeman et al., 2021).

To validate carbon credits, reductions must be additional, meaning they only occurred due to the carbon offset project. This is crucial for ensuring credits represent genuine environmental benefits. Ensuring additionality is challenging, as it requires proving that the carbon offset project would not have happened without carbon credit funding (Tyszkiewicz, 2022).

Carbon credit pricing is influenced by supply and demand dynamics within carbon markets. Prices fluctuate due to regulatory changes and the overall performance of carbon markets. Recent studies show that carbon credits in California's cap-and-trade program were valued at approximately \$13.67 per ton of CO₂ equivalent in 2011 (Ratnatunga & Jones et al., 2011). In 2011, the program was still in its early stages, and the pricing reflected the market's initial response to the new regulations. Over time, as the cap tightens and the number of available credits decreases, the price would typically be expected to increase. Prices also vary based on project type and quality, with credits from higher perceived environmental integrity projects, such as reforestation being valued higher. Certification by recognized standards like the Verified Carbon Standard or Gold Standard also enhances carbon credit value (Tyszkiewicz, 2022).

While carbon credit valuation has potential benefits, it also faces significant challenges and criticisms. A primary issue is baseline scenario manipulation. If baselines are set too high, projects can claim more credits without actually delivering proportional environmental benefits. This manipulation undermines the integrity of the carbon offset market (Ratnatunga & Jones et al., 2011). Project developers, sometimes without strict oversight from third-party verifiers, may set baselines too high to generate more carbon credits and increase profits. This results in "hot air" credits that do not correspond to actual emission reductions.

Another issue is over-crediting, where projects receive more credits than deserved due to statistical flaws (Dias & Henson et al., 2023). For example, using regional averages to establish baselines can lead to ecological fallacy, where generalized data misrepresents specific project conditions. This issue has been observed in California's forest carbon offsets program, where projects overestimate the baseline, leading to it being inflated—thus creating substantial over-crediting (Badgley & Freeman et al., 2021). This issue is largely due to a lack of standardization.

Double counting occurs when the same emissions reduction is claimed by multiple parties, such as the host country and the buyer of the carbon credit (Tyszkiewicz, 2022). This issue complicates the accurate accounting of emissions reductions. Along with this, variability in measurements introduces significant uncertainties that affect the reliability of credited reductions. These variabilities include things like measurement techniques and data quality (Opanda, 2024).

A critical challenge to address in carbon credit valuation is regulatory and verification framework. Without universal standards for carbon credit certification and valuation, discrepancies in credit quality and price arise. Different standards result in varying levels of environmental impact. If there are inconsistent regulations along with lack of rigorous third-party verification, this reduces the credibility of carbon credits (Favasuli & Sebastian, 2021). Lenient regulations can oversupply carbon credits, lowering prices and reducing funds for offset projects. This reduced funding means fewer resources for projects, resulting in fewer actual offsets compared to the credits bought.

Market volatility poses another challenge, as changes in regulatory policies and economic conditions can cause significant price fluctuations. Such volatility creates uncertainty for investors, hindering long-term planning and investment.

Addressing these challenges can enhance the carbon credit market's effectiveness in reducing greenhouse gas emissions and promoting sustainable development.

Discussion

The effectiveness of carbon markets in combating climate change is contingent upon addressing several critical challenges and implementing mechanisms to ensure transparency, credibility, and efficiency. This discussion will explore these challenges and propose an ideal carbon market model that fosters international cooperation and substantial emission reductions.

As discussed, carbon markets face several critical challenges, including baseline scenarios, over-crediting, lack of standardization, double counting, measurement uncertainty, and lacking regulatory & verification frameworks.

Baseline scenario manipulation occurs when projects set baselines too high, enabling projects to claim more credits without delivering proportional environmental benefits. To address this, rigorous and transparent methods for setting baselines must be implemented. Independent third-party verification can ensure that baselines are realistic and reflective of actual conditions.

Manipulation of the market can also lead to over-crediting. To combat over-crediting, standardized and precise measurement techniques should be adopted. Projects should be evaluated on a case-by-case basis to ensure that credit allocations are accurate and fair. In this case, the example set by the Paris Agreement should be followed, in which countries have to submit updated emission reduction plans every five years.

The absence of universal standards for carbon credit certification and valuation results in discontinuities regarding credit quality and price. This is significant as different standards lead to varying levels of environmental impact, complicating the assessment of the true value of carbon credits. Establishing universal standards, such as those set by the Verified Carbon Standard or

Gold Standard, would ensure consistency and credibility in carbon credits across markets. These standards are not limited to the United States; they are recognized and applied globally.

A prominent example is double-counting. Double counting occurs when the same emissions reduction is claimed by multiple parties, such as the host country and the buyer of the carbon credit. This issue complicates accurate accounting of emissions reductions and undermines the credibility of the carbon market. Combating this requires the use of blockchain technology, which prevents double counting by ensuring that each carbon credit is unique and traceable from creation to retirement.

An example is where each credit is logged on a blockchain with a unique identifier. As the company uses the credits to offset its emissions, the blockchain records the transaction, marking the credits as retired. Regulatory bodies and auditors can access the blockchain to confirm that the credits have not been double-counted, ensuring that the company's emission reductions are legitimate and accurately accounted for.

Variability in measurement techniques introduces significant uncertainties that affect the reliability of credited reductions. These differences in measurement methods and data quality lead to inconsistent and unreliable carbon credit valuations. Standardizing measurement protocols across projects will also reduce uncertainties. Utilizing advanced technologies like remote sensing and satellite monitoring can enhance the accuracy and consistency of emissions measurements.

Field surveys involve on-the-ground inspections and data collection by trained professionals. These surveys would be able to provide detailed, site-specific information about the project's impact on carbon emissions. By directly measuring factors, field surveys can validate actual emissions reductions. This allows for precise, localized data that can identify discrepancies or inaccuracies in the reported reductions, ensuring that the credits issued are based on real, verifiable outcomes.

Remote sensing can also be used in satellite imagery, drones, and other aerial technologies to monitor large areas of land or ocean. This method is particularly effective for projects spread over vast or inaccessible areas. This would provide up-to-date data, providing a bird's eye view to ensure that credits are based on thoroughly used data. These accurate verification methods would be able to properly help projects remain compliant with regulatory standards along with enhancing credibility of credits.

Inconsistent regulations and lack of rigorous third-party verification can lead to the oversupply of carbon credits, lowering prices and reducing funds available for offset projects. By strengthening regulatory frameworks and independent third-party verification, this will maintain the integrity of carbon credits. To do so, governments and international bodies would have to extensively collaborate to enforce rigorous regulatory frameworks.

An Ideal Carbon Market Model

First, an ideal carbon market would focus on establishing universal standards for carbon credit certification and valuation. To do so, recognized standards like the Verified Carbon

Standard or Gold Standard would have to be adopted globally to ensure consistency to establish credibility. The market should primarily prioritize projects with high environmental integrity, such as reforestation and renewable energy initiatives. These projects would not only be able to provide significant greenhouse gas reductions but also offer additional co-benefits, such as biodiversity conservation.

Strengthening regulations and transparency would address multiple issues. Being transparent in reporting baseline scenarios and verification processes is crucial for ensuring accountability. Clear and accessible reporting will help stakeholders, including investors and regulators, make informed decisions and ensure that carbon credits are credible and effective. Furthermore, by establishing the blockchain technology, auditors can access the blockchain to verify the authenticity and status of credits, ensuring proper compliance with regulatory standards.

By constantly ensuring proper compliance with regulations, this consistency will prove beneficial, as uniform standards ensure that carbon credits are issued, measured, and verified consistently, making them easier to compare and trade across different regions. This ensures that all credits undergo the same level of scrutiny, reducing the risk of fraud and ensuring emissions reductions are real and verifiable.

Advanced verification methods, such as field surveys and remote sensing, are essential tools for ensuring that reported reductions in greenhouse gas emissions are credible and quantifiable. This would provide accurate, real-time data on the performance of carbon offset projects, helping to verify that the claimed reductions are genuinely achieved.

Conclusion

By examining the mechanisms, economics, and implementation strategies of carbon markets, this paper highlights the importance of rigorous standards, transparency, and advanced verification to enhance the effectiveness and credibility of these markets in reducing greenhouse gas emissions. Establishing a robust and globally consistent carbon market model is crucial for fostering international cooperation and ensuring long-term climate resilience.

Works Cited

- "Carbon Trading Continues: What's Wrong with the Voluntary Market." Global Witness. Last modified May 24, 2022. Accessed June 13, 2024.
<https://www.globalwitness.org/en/campaigns/greenwashing/carbon-trading-continues-whats-wrong-with-the-voluntary-market/>
- "Climate Impacts on Agriculture and Food Supply." Chicago Department of Climate Change. Accessed April 21, 2024.
<https://climatechange.chicago.gov/climate-impacts/climate-impacts-agriculture-and-food-supply#:~:text=Moderate%20warming%20and%20more%20carbon,quality%20of%20their%20food%20supply.>
- "Climate Impacts." Union of Concerned Scientists. Accessed April 20, 2024.
<https://www.ucsusa.org/climate/impacts#:~:text=Climate%20change%20is%20also%20making,to%20once%20every%2016%20years>
- "Intergovernmental Panel on Climate Change (IPCC)." European Commission. Accessed June 24, 2024.
[https://research-and-innovation.ec.europa.eu/research-area/environment/climate-change-science/intergovernmental-panel-climate-change-ipcc_en#:~:text=The%20Intergovernmental%20Panel%20on%20Climate%20Change%20\(IPCC\)%2C%20consisting%20of,the%20work%20of%20thousands%20of](https://research-and-innovation.ec.europa.eu/research-area/environment/climate-change-science/intergovernmental-panel-climate-change-ipcc_en#:~:text=The%20Intergovernmental%20Panel%20on%20Climate%20Change%20(IPCC)%2C%20consisting%20of,the%20work%20of%20thousands%20of)
- "The Paris Agreement." United Nations. Accessed April 19, 2024.
<https://www.un.org/en/climatechange/paris-agreement>
- "The Ultimate Guide to Understanding Carbon Credits." CarbonCredits.com. Accessed April 17, 2024. <https://carboncredits.com/the-ultimate-guide-to-understanding-carbon-credits/>
- "What Are Carbon Markets and Why Are They Important?" United Nations Development Programme (UNDP) Climate Promise. Last modified May 18, 2024. Accessed May 22, 2024.
<https://climatepromise.undp.org/news-and-stories/what-are-carbon-markets-and-why-are-they-important>
- "What Is the Voluntary Carbon Market?" CarbonCredits.com. Accessed June 11, 2024.
<https://carboncredits.com/what-is-the-voluntary-carbon-market/>
- Alsaifi, Khaled, Marwa Elnahass, and Aly Salama. "Market Responses to Firms' Voluntary Carbon Disclosure: Empirical Evidence from the United Kingdom." *Journal of Cleaner Production* 262 (2020): 121377. Accessed June 18, 2024.
<https://doi.org/10.1016/j.jclepro.2020.121377>
- Badgley, Grayson, Jeremy Freeman, Joseph J. Hamman, Barbara Haya, Anna T. Trugman, William R. L. Anderegg, and Danny Cullenward. "Systematic Over-Crediting in California's Forest Carbon Offsets Program." *Global Change Biology* 27, no. 24 (2021): 6435-6450. Accessed June 20, 2024. <https://doi.org/10.1111/gcb.15943>
- Bassetti, Francesco. "Success or Failure? The Kyoto Protocol's Troubled Legacy." *Foresight*. Last modified December 8, 2022. Accessed June 18, 2024.
<https://www.foresightmcc.com/success-or-failure-the-kyoto-protocols-troubled-legacy>
- Cambridge Institute for Sustainability Leadership "What is COP?" University of Cambridge. Accessed June 24, 2024.
<https://www.cisl.cam.ac.uk/cop-climate-change-conference#:~:text=COP%20stands%20for%20Conference%20of,making%20body%20of%20the%20UNFCCC>
- Cassidy, Bill. "Cassidy Leads Republican Senate Opposition to a Carbon Tax." Press Release. Last modified October 26, 2023. Accessed June 16, 2024.

- <https://www.cassidy.senate.gov/newsroom/press-releases/cassidy-leads-republican-senate-opposition-to-a-carbon-tax/#:~:text=%E2%80%9CIncreasing%20the%20cost%20of%20energy,Cassidy>
- Dawes, Allegra. "What's Plaguing Voluntary Carbon Markets?" CSIS. Last modified February 2, 2024. Accessed June 10, 2024. <https://www.csis.org/analysis/whats-plaguing-voluntary-carbon-markets>
- Dias, Ana Karla Nobre, Alex Henson, Matt Lavelle, and Kirti Ramesh. "BeZero's Carbon Risk Factor Series: Over-Crediting." BeZero. Last modified December 14, 2023. Accessed June 20, 2024. <https://www.bezero.com/carbon-risk-factor-series-over-crediting>
- Eden, Alexander, Charlotte Unger, William Acworth, Kristian Wilkening, and Constanze Haug. "Benefits of Emissions Trading: Taking Stock of the Impacts of Emissions Trading Systems Worldwide." International Carbon Action Partnership. Last modified August 2018. Accessed May 22, 2024. https://icapcarbonaction.com/system/files/document/benefits-of-ets_updated-august-2018.pdf
- Favasuli, Silvia, and Vandana Sebastian. "Voluntary carbon markets: how they work, how they're priced and who's involved." S&P Global Commodity Insights. Last modified June 10, 2021. Accessed April 22, 2024. <https://www.spglobal.com/commodityinsights/en/market-insights/blogs/energy-transition/061021-voluntary-carbon-markets-pricing-participants-trading-corsia-credits>
- Goulder, Lawrence H., and Andrew Schein. "Carbon Taxes vs. Cap and Trade: A Critical Review." National Bureau of Economic Research Working Paper No. 19338. Last modified August 2013. Accessed June 11, 2024. <https://www.nber.org/papers/w19338>
- Ho, Ben. "Time to Give Up on a Carbon Tax?" State of the Planet, Columbia Climate School. Last modified April 26, 2021. Accessed June 16, 2024. <https://blogs.ei.columbia.edu/2021/04/26/time-to-give-up-on-a-carbon-tax/>
- Kreibich, Nicolas, and Lukas Hermwille. "Caught in Between: Credibility and Feasibility of the Voluntary Carbon Market Post-2020." Taylor & Francis Online Climate Policy 21, no. 7 (2021). Last modified July 7, 2021. Accessed April 22, 2024. <https://doi.org/10.1080/14693062.2021.1948384>
- Lam, Karen, Jenny Wong, and Gilly Hutchinson. "Final Regulations Published for the China Certified Emission Reduction Scheme." Linklaters. Last modified October 30, 2023. Accessed June 20, 2024. <https://www.linklaters.com/en/insights/publications/2023/october/final-regulations-published-for-china-certified-emission-reduction-scheme>
- Lindsey, Rebecca. "Climate Change: Global Sea Level." Climate.gov. Last modified April 19, 2022. Accessed April 20, 2024. <https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level#:~:text=What's%20causing%20sea%20level%20to,expanding%20as%20the%20water%20warms>
- Lo, Alex Y. "Challenges to the development of carbon markets in China." Taylor & Francis Online Climate Policy 16, no. 1 (2016): 109-124. Last modified January 22, 2015. Accessed June 24, 2024. <https://doi.org/10.1080/14693062.2014.991907>
- Mendelsohn, Robert, Robert Litan, and John Fleming. "How to Repair the World's Broken Carbon Offset Markets." Yale Environment 360. Last modified November 18, 2021. Accessed June 18, 2024. <https://e360.yale.edu/features/how-to-repair-the-worlds-broken-carbon-offset-markets>

- Metcalf, Gilbert E. "Designing a Carbon Tax to Reduce U.S. Greenhouse Gas Emissions." National Bureau of Economic Research Working Paper No. 14375. Last modified October 2008. Accessed June 14, 2024. <https://www.nber.org/papers/w14375>
- Nordhaus, Ted, and Michael Shellenberger. "The Flawed Logic of the Cap-and-Trade Debate." Yale Environment 360. Last modified May 19, 2009. Accessed June 24, 2024. https://e360.yale.edu/features/the_flawed_logic_of_the_cap-and-trade_debate
- Omolere, Mitota P. "Explainer: What Is a Carbon Tax, Pros and Cons, and Implementation Around the World." Earth.org. Last modified January 9, 2024. Accessed June 15, 2024. <https://www.earth.org/explainer-what-is-a-carbon-tax-pros-and-cons-and-implementation-around-the-world>
- Opanda, Samson. "Carbon Credit Pricing Chart: Updated 2023." 8 BillionTrees.com. Last modified March 15, 2024. Accessed June 18, 2024. <https://8billiontrees.com/carbon-offsets-credits/new-buyers-market-guide/carbon-credit-pricing/#:~:text=There%20are%20many%20ways%20to,location%2C%20and%20other%20determining%20factors>
- Rathnunge, Janek, Stewart Jones, and Kashi R. Balachandran. "The Valuation and Reporting of Organizational Capability in Carbon Emissions Management." Accounting Horizons 25, no. 1 (2011): 127-147. Accessed June 20, 2024. <https://doi.org/10.2308/acch.2011.25.1.127>
- Song, Lisa. "Cap and Trade Is Supposed to Solve Climate Change, but Oil and Gas Company Emissions Are Up." ProPublica. Last modified November 15, 2019. Accessed June 24, 2024. <https://www.propublica.org/article/cap-and-trade-is-supposed-to-solve-climate-change-but-oil-and-gas-company-emissions-are-up>
- Sumner, Jenny, Lori Bird, and Hillary Smith. "Carbon Taxes: A Review of Experience and Policy Design Considerations." National Renewable Energy Laboratory. Last modified December 2009. Accessed May 23, 2024. https://digital.library.unt.edu/ark:/67531/metadc929933/m2/1/high_res_d/970341.pdf
- Trouwloon, Danick, Charlotte Streck, Thiago Chagas, and Glenpherd Martinus. "Understanding the Use of Carbon Credits by Companies: A Review of the Defining Elements of Corporate Climate Claims." Global Challenges 7, no. 4 (April 2023): 2200158. Last modified February 25, 2023. Accessed June 18, 2024. <https://doi.org/10.1002/gch2.202200158>
- Tyszkiewicz, Elisabeth. "Decarbonizing Economics through Carbon Pricing - Is there an ideal policy mix?" Master's work project, Nova School of Business and Economics. Last modified March 10 2022. Accessed June 20, 2024. https://run.unl.pt/bitstream/10362/152706/1/48702_elisabethtyszkiewicz.pdf
- University of Exeter and Stanford Doerr School of Sustainability. "Global Carbon Emissions from Fossil Fuels Reached Record High in 2023." Stanford Sustainability. Last modified December 5, 2023. Accessed June 20, 2024. <https://sustainability.stanford.edu/news/global-carbon-emissions-fossil-fuels-reached-record-high-2023>
- Xiaoying, You. "As China's carbon market turns two, how has it performed?" Dialogue Earth. Last modified August 31, 2023. Accessed May 22, 2024. <https://dialogue.earth/en/climate/china-carbon-market-turns-two-how-has-it-performed/#:~:text=Chen%20says%20that%2C%20currently%2C%20the,and%20economy%20are%20still%20growing>
- Zhou, Kaile, and Yiwen Li. "Carbon Finance and Carbon Market in China: Progress and Challenges." Journal of Cleaner Production 214 (2019): 536-549. Accessed May 22, 2024. <https://dpl6hyzg28thp.cloudfront.net/media/mack2.pdf>

Simulating the Effects of Weighted Dead Hangs on Spacing and Stretch of Glenohumeral Joints By Ian Lee

Summary

Investigating how vertical force from weighted dead hangs affects scapula-humerus spacing and strains the glenohumeral ligament to optimize shoulder rehabilitation.

Abstract

Shoulder pain is a prevalent issue, often hindering daily activities such as sleeping or reaching overhead. Dead hangs, an exercise where one hangs from a bar with arms fully extended, are commonly recommended by physical therapists for their benefits in decompressing the shoulder joint, enhancing mobility, and reducing the risk of impingement by increasing joint space and circulating synovial fluid. Despite their widespread use, the specific biomechanics of how the vertical force from a dead hang influences the spacing between the scapula and humerus, and the resulting strain on the glenohumeral ligament, have not been fully studied.

This study aims to fill this knowledge gap by investigating the effects of vertical forces, proportional to body weight, on the spacing between the scapula and humerus during weighted dead hangs. A shoulder joint model was constructed using plastic bone models and rubber bands to simulate the glenohumeral ligaments. Weights were incrementally added to the model to measure changes in joint spacing and ligament strain. The tensile strength of the rubber bands was calculated to provide a basis for comparison with human ligaments.

Our findings reveal a clear relationship between the applied force and the increase in joint spacing, with the rubber bands stretching proportionally to the added weights. This data suggests that the tensile strength and elasticity of human ligaments can be modeled to predict the effects of dead hangs on shoulder joint mechanics. The study provides a preliminary model for physical therapists and orthopedic surgeons, helping them optimize the use of dead hangs in shoulder rehabilitation.

Key terms: *glenohumeral joint, glenohumeral ligaments, shoulder rehabilitation, shoulder pain*

Introduction

Shoulder pain is a prevalent issue, affecting 18-26% of adults at any given time, making it one of the most common regional pain syndromes according to population surveys (Linaker and Walker-bone). Shoulder pain may impact simple day-to-day activities such as sleeping at night or reaching for an object over one's head. Dead hangs are commonly recommended by physical therapists due to their effectiveness in treating shoulder pain. Dead hangs have many benefits, such as decompressing the shoulder joint, enhancing mobility, and reducing the risk of impingement by increasing joint space and circulating synovial fluid, which lubricates the joints. To perform a dead hang, one must hang from a bar with both arms fully extended and with both feet off the ground, exerting a vertical force pulling upwards on the shoulder joints.

Despite the widespread use of dead hangs for shoulder rehabilitation, the specific biomechanics of how the force of a dead hang affects the spacing between the humerus and the scapula and how it stretches the glenohumeral ligament has yet to be fully understood. Understanding how the dead hang affects the shoulder joint is crucial for optimizing this rehabilitation method in order to further increase its effectiveness and prevent possible injury.

To face this knowledge gap, this study investigates the effects of the vertical force, proportional to body weight, on the spacing between the scapula and the humerus during weighted dead hangs. This study measures the strain experienced by the glenohumeral ligament under these conditions to compare how different forces affect the joint spacing between the humerus and the scapula through the construction of a shoulder joint model and adding weights. This data will provide valuable insights into the biomechanics of dead hangs and in understanding how to further increase its effectiveness and prevent possible injury.

Methodology

Materials

In order to test the effects of the vertical force from dead hangs on the joint spacing between the humerus and the scapula, a shoulder joint model was created. The materials used to represent the bones were plastic bone models of the humerus and scapula with dimensions and weights 13.7 x 2.68 x 1.54 inches and 7.05 ounces and 6.14 x 4.72 x 2.13 inches and 3.74 ounces, respectively (anatomywarehouse.com). The glenohumeral ligaments were represented by size #33 rubber bands with dimensions 3½ x ¼ inches (amazon.com). The bar used for dead hangs was a doorway pull-up bar (amazon.com). Lastly, nine 50-gram lab weights attached to a metal hook that was also 50 grams were used to provide differing vertical forces from the dead hangs.

Construction of Model

The model bone of the humerus, facing vertically upwards, was put into the glenoid cavity of the scapula through the use of one rubber band to represent the glenohumeral ligaments. The rubber band was wrapped once around the head of the humerus to get a good hold and then wrapped around the acromion of the scapula. This leads to an “X” shaped configuration. The ends of the rubber bands were all then super glued onto the bones. The model was then weighed at 0.68 lbs.

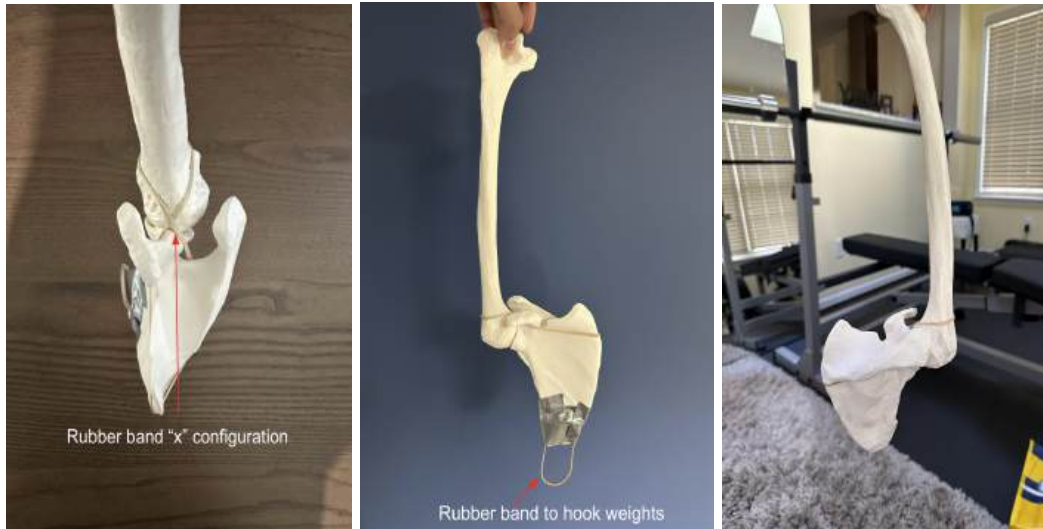


Figure 1: Depiction of model with rubber band configuration with no weight applied.

Experimental Set-up

The distal end of the humerus is then duct taped to the bar in a circular path with around ten inches of tape, and tightly enough to allow for the least amount of pliancy. The 50 grams hook is hooked onto the rubber band that is duct taped to the inferior angle of the scapula in the same way as mentioned before, weights are added in increments of 50 grams. The holes in the weights are placed onto the shank so that the weights slide down to the bottom. Lastly, measure the tensile strength of the rubber band to make applicable results. First, find the cross-sectional area of the rubber band. Then, find the amount of weight the rubber band can hold by clamping the top and setting dumbbells in the rubber band. Gradually go up in weight until it snaps. To find the tensile strength, do the amount of weight the rubber band can hold divided by the cross-sectional area.

Data Collection

After the construction of the model is complete, the data must be recorded. The initial length of the rubber band without any added weight was measured. In order to measure the length of this, measure starting from the "X" shaped cross of the rubber band, to the tip of the scapula (Figure 3). This is because as more weight is added, the "X" shaped section of the rubber band remains securely positioned on top of the humerus, without moving or stretching. Only the portions of the rubber band extending beyond the "X" shape experience any stretching. Then, the 50-gram hook was added. Wait around 20 seconds due to the shaking and bouncing of the rubber band within the model after adding weight. After the model is static, the length of the rubber band was measured in the same way as mentioned before. Then add a 50-gram weight and measure again. This procedure was repeated 8 more times until all 9 weights were placed onto the hook for a total of 500 grams.



Figure 2: Shows the area on the model that is measured by the ruler.

Rubber Band Assessment

Find the stiffness of the rubber band. Just as before, clamp the top of the rubber band, but instead of finding the maximum force that the rubber band can handle, use the 50-gram hook, add weights in increments of 50 grams, and measure the length of the rubber band after each weight increment. The difference between this and the experiment is that this is testing the stiffness of the rubber band under no special configuration. After the data has been collected, graph the data points and find the slope through a trendline to find the stiffness of the rubber band.

Data Analysis

Tensile Strength:

1. width \times thickness \times two = cross-sectional area
2. $\frac{\text{maximum weight rubber band can hold}}{\text{cross-sectional area}} = \text{tensile strength}$

Stiffness:

1. $F=kx$
2. $F = \text{weight applied in newtons}$
3. $x = \text{length of rubber band in meters}$
4. $k = \text{stiffness}$
5. $\frac{F}{x} = k$
6. $\frac{\text{average weight applied}}{\text{average length of rubber band}} = \text{stiffness}$

Results

The table shows the raw results, showing the full lengths of the rubber band at different weights applied. Stretch was calculated by subtracting 0.597 cm from all average lengths to find the average stretch of each weight applied. At the highest weight point, the stretch was calculated to be 2.12 cm.

Weight applied (g)	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Average Length and Standard Deviation (cm)	Average Stretch (cm)
0	0.600	0.610	0.580	0.597 ± 0.0153	0
50	0.820	0.770	0.810	0.800 ± 0.0265	0.203
100	1.03	0.970	1.04	1.01 ± 0.0379	0.413
150	1.25	1.11	1.28	1.21 ± 0.0907	0.613
200	1.56	1.37	1.41	1.45 ± 0.100	0.853
250	1.75	1.55	1.63	1.64 ± 0.101	1.04
300	1.91	1.88	1.85	1.88 ± 0.0300	1.28
350	2.13	2.04	2.00	2.06 ± 0.0666	1.46
400	2.40	2.24	2.36	2.33 ± 0.0833	1.73
450	2.51	2.46	2.58	2.52 ± 0.0603	1.92
500	2.75	2.66	2.75	2.72 ± 0.0520	2.12

Figure 3: Summarizes the results of the experiment. The average length at 0 grams applied was 0.597 cm and at 500 grams applied was 2.72 cm.

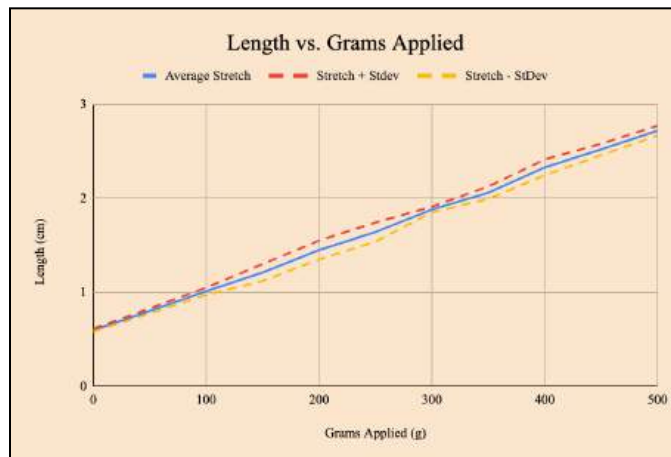


Figure 4: Graph showing the linear relationship between length vs grams applied along with the length \pm standard deviation.

Tensile Strength of the Rubber Band

With the width and thickness being 1.905×10^{-2} meters and 1.016×10^{-3} meters respectively, the cross-sectional area was 3.872×10^{-5} square meters. The rubber band could withstand at most 66.7233 newtons. Therefore, the tensile strength was 1.723 MPa.

Stiffness of the Rubber band

The length vs force applied was tested when the rubber band was not in the configuration of the “X” shape within the model. The rubber band in this case was just clamped at the top and a weight was hooked onto the bottom of the band. Through Hooke’s law $F=kx$, the k value was graphed and found based on the data in figure. The k value was found to be 78.6 N/m.

Length [m]	Newtons Applied [N]
0.0813	0.000
0.0864	0.490
0.0889	0.981
0.0914	1.47
0.0965	1.96
0.102	2.45
0.107	2.94
0.119	3.43
0.124	3.92
0.132	4.41
0.142	4.90

Figure 5: Table of length in meters vs newtons applied of the rubber band with no “X” configuration

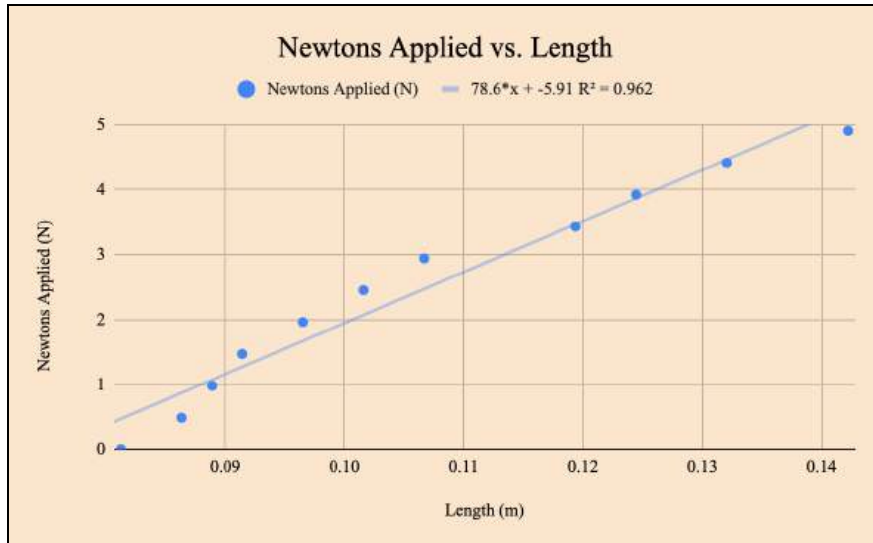


Figure 6: Illustrates the relationship between the newtons applied and the length of the rubber band in meters with an R^2 value of 0.962.



Figure 7: Pictures of the model under different weights applied

Discussions

The results clearly show that as the weight applied increases, the distance between the humerus and the scapula increases as well in a linear fashion. Understanding this relationship provides valuable insights into the biomechanical benefits of dead hangs, offering a scientific basis for their use in physical therapy and rehabilitation programs.

All the ligaments within the glenohumeral joint are the Superior Glenohumeral Ligament, Middle Glenohumeral Ligament, Inferior Glenohumeral Ligament, and Coracohumeral Ligament. Their average tensile strengths are 10 MPa, 13.3 MPa, 15 MPa, and 12.5 MPa, respectively. In total their tensile strength equates to 50.8 MPa, meaning that the tensile strength of the shoulder ligaments is 28.48 times greater than the tensile strength of the rubber band used to represent them. With this knowledge, if the weight applied by our testing is multiplied by 28.48, then one will achieve the same distance moved between their scapula and humerus as shown in the data. The findings could help physical therapists optimize rehabilitation protocols,

ensuring that patients receive the most effective treatments for shoulder pain and mobility issues. The data can help to apply the most effective weights to increase joint mobility and decrease pain.

The findings from this study have significant future implications. Through further research, the data could further influence medical practices, leading to more refined rehabilitation programs for shoulder pain. The study suggests that carefully monitored weighted dead hangs could be a powerful tool in improving shoulder joint health and reducing pain.

While the model used in this study provided valuable data, it has several limitations. The primary limitation is the use of rubber bands to simulate the glenohumeral ligaments. Rubber bands do not perfectly replicate the elastic and tensile properties of human ligaments, which may affect the accuracy of the results. Additionally, the static nature of the model does not take into account movements and muscle activity present in humans. To strengthen the study, future research could incorporate tendon grafts and synthetic ligaments into the model to simulate the natural elasticity and tensile properties of human ligaments more accurately. Looking even further, computer-generated simulations may provide even more accurate results.

In conclusion, while the study presents promising data on the benefits of dead hangs for shoulder joint decompression, it emphasizes the need for continued research to translate them into practical medical applications. By addressing the limitations and building on this foundational work, future studies can significantly enhance our understanding and treatment of shoulder pain.

Works Cited

- Bang, Yun-Sic et al. "The value of the glenohumeral joint cross-sectional area as a morphological parameter of glenohumeral osteoarthritis." *Medicine* vol. 101,47 (2022): e31424. doi:10.1097/MD.00000000000031424
- Bigliani, L U et al. "Tensile properties of the inferior glenohumeral ligament." *Journal of orthopaedic research : official publication of the Orthopaedic Research Society* vol. 10,2 (1992): 187-97. doi:10.1002/jor.1100100205
- Boardman, N D et al. "Tensile properties of the superior glenohumeral and coracohumeral ligaments." *Journal of shoulder and elbow surgery* vol. 5,4 (1996): 249-54. doi:10.1016/s1058-2746(96)80050-4
- Chang LR, Anand P, Varacallo M. *Anatomy, Shoulder and Upper Limb, Glenohumeral Joint*. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK537018/>
- Dan Karlsson, Bo Peterson, Towards a model for force predictions in the human shoulder, *Journal of Biomechanics*, Volume 25, Issue 2, 1992, Pages 189-199, ISSN 0021-9290, [https://doi.org/10.1016/0021-9290\(92\)90275-6](https://doi.org/10.1016/0021-9290(92)90275-6). (<https://www.sciencedirect.com/science/article/pii/0021929092902756>)
- Linaker, Catherine H, and Karen Walker-Bone. "Shoulder disorders and occupation." *Best practice & research. Clinical rheumatology* vol. 29,3 (2015): 405-23. doi:10.1016/j.berh.2015.04.001
- López-Rivera, Eva & González-Badillo, Juan. (2019). The Effects of a Weighted Dead-Hang training program on Grip Strength and Endurance in Expert climbers with different levels of strength.
- Singh, Rajani. "Surgical Anatomy of the Glenoid Cavity and Its Use in Shoulder Arthroplasty Among the North Indian Population." *Cureus* vol. 12,12 e11940. 6 Dec. 2020, doi:10.7759/cureus.11940
- Zhu, Xianzhong, and Fuguo Huang. *Zhongguo xiu fu chong jian wai ke za zhi = Zhongguo xiufu chongjian waikewaxue zazhi = Chinese journal of reparative and reconstructive surgery* vol. 21,11 (2007): 1175-8.

A Novel Approach to Treat Parkinson's Disease By Emily Zhang

Abstract:

Parkinson's Disease is a common neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the brain. This review paper discusses the current clinical treatments and medication that address this disorder and the downsides associated with each. An analysis of different types of stem cell therapy and their potential challenges is also discussed. Finally, the review paper presents a recent cell-free paracrine approach known as mesenchymal cell-derived extracellular vesicles aiming to address barriers posed by previous approaches and its prospects of delivering maximal patient outcomes, and symptom recovery, with fewer risks.

Introduction:

Parkinson's Disease (PD) is a neurodegenerative disorder that affects several regions of the brain, particularly the basal ganglia and the substantia nigra. Nerve cells residing in the substantia nigra are responsible for producing the neurotransmitter dopamine, which controls movement and balance. Nerve cells in patients with Parkinson's disease weaken or die gradually, leading to symptoms such as tremors, stiffness in the body, speech difficulties, trouble with balance, and others [5]. Parkinson's is the second most common neurodegenerative disease with about 90,000 people in the U.S. diagnosed every year [8]. Although not fully understood, the causes of Parkinson's disease involve genetic and environmental factors. The biomarker used to diagnose PD is the abnormal clumps of protein alpha-synuclein which hinders communications between A9-type midbrain dopaminergic (mDA) neurons (mDANs) [15].

Current clinical treatments:

PD isn't currently considered curable, however, there are many treatments aimed at mitigating the symptoms caused by the disease. A widely employed, non-invasive treatment is the use of dopamine precursors such as levodopa as medication to replenish the lack of dopamine production in damaged neurons [6]. Drugs that influence other neurotransmitters and antidepressants are also used to alleviate tremor issues and PD-related depression respectively. However, these medications come with a major problem: patients' bodies will develop resistance and the efficacy can decrease dramatically over time. Another possible treatment option is Deep Brain Stimulation, which implants electrodes to generate electrical impulses that can modulate neurotransmitters within the brain. Yet, this treatment is invasive, poses risks of stroke and infection, and isn't suitable for everyone. For some, it could lead to memory lapses, mood changes, or loss of coordination without much improvement of current symptoms [13].

Stem cell therapy:

In recent decades, scientists have been working on stem cell therapy as a potential

physiological treatment that could lead to promising results in PD patients, though this approach is still in the clinical trials phase and has not yet been approved by the FDA [16]. Research has been centered around finding the best source of stem cells for maximal efficacy and minimal harm.

Stem cells are a special type of cells that can regenerate, make duplicates of themselves, and develop into different tissues varied by their location in a process known as differentiation [10]. The earliest use of stem cells in a clinical trial was derived from fetal ventral mesencephalic (FVM) tissues dissected from aborted embryos [12]. Many trials have shown patients who received the treatment experienced improving symptoms and signs of recovery [14]. Despite this, using aborted fetuses as a source raises ethical challenges and thus cannot be implemented long-term for widespread use.

This led to scientists exploring other sources of stem cells, each with unique potential and shortcomings. The source could be either allogeneic (from a different source other than the patient) or autologous (extracted from the patient). Both human embryonic stem cells (hESC) derived from human blastocysts and human induced pluripotent stem cells (hiPSC)-from either a random donor or an HLA-matched donor-are prone to mutations and immunological issues in which the host immune cells might attack the graft (inserted stem cells). Specifically, the host immune cells might not recognize the major histocompatibility complex (MHC) and cause an attack [7]. There are two ways to solve this problem: tissue typing and immunosuppression. Tissue typing involves finding a compatible match that has similar MHC and self-antigens as the host. However finding a suitable match could be difficult and if matched, partial mismatch is still possible, leading to rejection of the graft. Alternatively, immunosuppression is a method to suppress the immune system ability of the host, hoping for less of an immune response or rejection upon insertion of the graft. This process, nevertheless, is highly dangerous and unpredictable because a decreased function of the host immune system drastically lowers the body's ability to fight off pathogens, as such, even a common cold poses the potential to endanger the patient's life. With that said, autologous grafts (involving transforming patients' somatic cells into hiPSC) would therefore be the best option among the rest in theory. That may be the case, yet its lengthy and costly process hinders widespread use.

MSC-EV:

Although stem cell therapy holds a promising future to fundamentally improve PD symptoms, immunological and practicality barriers exist, and thus some scientists directed their focus to cell-free approaches [4]. The particular approach this review paper will focus on is the usage of mesenchymal stem cell-derived extracellular vesicles (MSC-EVs).

Mesenchymal stem cells are isolated from tissues like bone marrow, umbilical cord, and adipose tissue [9] and have been used in animal trials to demonstrate their role. In a study [1], MSC was infused into baboons (nonhuman primates) to assess its regenerative capacity. The results showed therapeutic effects upon infusion, but the percentage of differentiated, detected MSC in their analysis was only 0.1% - 2.7%. In another study [2], researchers infused

adipose-derived mesenchymal stem cells into transgenic mice (modeling ALS-Amyotrophic lateral sclerosis) to administer its effect. The results revealed that treated mice showed better motor performance, a higher number of motor neurons in the lumbar, and a restricted number of undifferentiated mesenchymal stem cells (marked by green fluorescent protein labels) in the spinal cord. After careful examination of the spinal cord tissue factor, the team found significantly higher neurotrophic factor (GDNF) and basic fibroblast growth factor (bFGF) in the treated mice group. Since adipose-derived mesenchymal stem cells(ASC) are only capable of producing bFGF and not GDNF, the team concluded ASC's neuroprotective ability through secretomes. This means that mesenchymal stem cells use paracrine pathways to deliver their content to target cells.

Additionally, due to MSC-EVs small size, they can pass through the blood-brain barrier (BBB), targeting neurodegenerative diseases specifically. EVs successfully reached the substantia nigra through the BBB, reduced substantia nigra dopaminergic neuron loss and apoptosis (cell death), and upregulated the levels of dopamine in the sample in a recent study [3].

MSC-EV has therefore shown promising leads to become a cell-free treatment for PD without risks of immunological and tumorigenic issues. Granted, some challenges still need to be addressed before MSC-EVs are widely adopted in clinical practice. For example, standardization of production methods, quantification and characterization, and monitoring potential side effects (since this approach is still relatively new) in different individuals [11].

Conclusion:

In conclusion, Parkinson's disease still faces challenges in treatment and management. While traditional neurotransmitter-modulating medication offers relief in symptoms, the results diminish with time. Stem cell therapy has shown efficacy in different research, but ethical concerns, immunological risks, and procedural complexities prevent it from widespread clinical use currently. Alternatively, MSC-EVs address many issues associated with previous approaches and with additional research, standardization, and monitoring could potentially improve the quality of life for many PD patients.

Works Cited

1. Bhandari, T. (2019, December 10). Why doesn't deep-brain stimulation work for everyone? Washington University School of Medicine in St. Louis.
<https://medicine.wustl.edu/news/new-clues-to-why-deep-brain-stimulation-doesnt-always-help-patients/#:~:text=Such%20stimulators%20are%20designed%20to>
2. Cha, Y., Park, T.-Y., Leblanc, P., & Kim, K.-S. (2023). Current Status and Future Perspectives on Stem Cell-Based Therapies for Parkinson's Disease. *Journal of Movement Disorders*.
<https://doi.org/10.14802/jmd.22141>
3. Chen, H.-X., Liang, F.-C., Gu, P., Xu, B.-L., Xu, H.-J., Wang, W.-T., Hou, J.-Y., Xie, D.-X., Chai, X.-Q., & An, S.-J. (2020). Exosomes derived from mesenchymal stem cells repair a Parkinson's disease model by inducing autophagy. *Cell Death & Disease*, 11(4), 1–17.
<https://doi.org/10.1038/s41419-020-2473-5>
4. Devine, S. M., Cobbs, C., Jennings, M., Bartholomew, A., & Hoffman, R. (2003). Mesenchymal stem cells distribute to a wide range of tissues following systemic infusion into nonhuman primates. *Blood*, 101(8), 2999–3001.
<https://doi.org/10.1182/blood-2002-06-1830>
5. Gilbert, R. (2024, July 3). Stem Cell Therapy in Parkinson's Disease | APDA. American Parkinson Disease Association.
<https://www.apdaparkinson.org/article/understanding-stem-cell-therapy-in-parkinsons-disease-treatment/#:~:text=There%20is%20no%20FDA%20approved>
6. Gowen, A. (2020, March 11). Mesenchymal Stem Cell-Derived Extracellular Vesicles: Challenges in Clinical Applications. *Frontiers*.
<https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcel.1.2020.00149/full>
7. K, A. (2024). - YouTube. [youtu.be. https://youtu.be/0y1UOq7EIMg?si=Vi0t4broUrkw-IRt](https://youtu.be/0y1UOq7EIMg?si=Vi0t4broUrkw-IRt)
8. Kou, M., Huang, L., Yang, J., Chiang, Z., Chen, S., Liu, J., Guo, L., Zhang, X., Zhou, X., Xu, X., Yan, X., Wang, Y., Zhang, J., Xu, A., Tse, H., & Lian, Q. (2022). Mesenchymal stem cell-derived extracellular vesicles for immunomodulation and regeneration: a next generation therapeutic tool? *Cell Death & Disease*, 13(7).
<https://doi.org/10.1038/s41419-022-05034-x>
9. Marconi, S. (2013, September 17). Systemic treatment with adipose-derived mesenchymal stem cells ameliorates clinical and pathological features in the amyotrophic lateral sclerosis murine model. *ScienceDirect*.
<https://www.sciencedirect.com/science/article/abs/pii/S0306452213004545>
10. Mayo Clinic Staff. (2022, March 19). Stem cells: What they are and what they do. Mayo Clinic; Mayo Foundation for Medical Education and Research.
<https://www.mayoclinic.org/tests-procedures/bone-marrow-transplant/in-depth/stem-cells/art-20048117>

11. National Institute of Neurological Disorders and Stroke. (2023). Parkinson's Disease | National Institute of Neurological Disorders and Stroke. [Www.ninds.nih.gov](http://www.ninds.nih.gov).
<https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease>
12. Park, T.-Y., Jeon, J., Cha, Y., & Kim, K.-S. (2024). Past, present, and future of cell replacement therapy for parkinson's disease: a novel emphasis on host immune responses. *Cell Research*, 1–14. <https://doi.org/10.1038/s41422-024-00971-y>
13. Parkinson's disease: MedlinePlus Genetics. (2012, May 1). [Medlineplus.gov](http://medlineplus.gov).
<https://medlineplus.gov/genetics/condition/parkinsons-disease/#causes>
14. Parkinson's Foundation. (2024). Statistics. Parkinson's Foundation; Parkinson's Foundation.
<https://www.parkinson.org/Understanding-Parkinsons/Statistics>
15. Tracking the spread of Parkinson's proteins from gut to brain. (2019, July 22). National Institutes of Health (NIH).
<https://www.nih.gov/news-events/nih-research-matters/tracking-spread-parkinsons-proteins-gut-brain>
16. Turano, E., Iliara Scambi, Federica Virla, Bonetti, B., & Mariotti, R. (2023). Extracellular Vesicles from Mesenchymal Stem Cells: Towards Novel Therapeutic Strategies for Neurodegenerative Diseases. *International Journal of Molecular Sciences*, 24(3), 2917–2917. <https://doi.org/10.3390/ijms24032917>

Isolation of S. Mutans Bacteriophage and Investigation of S. Mutans Metabolic Versatility Across Media By Ang L. Li

Abstract

Dental caries is one of most pervasive chronic infections in the world. Despite the emergence of many therapies targeted for dental caries, this remains as a global concern. The cariogenicity of tooth enamel has recently been found to be caused by *Streptococcus Mutans*; an anaerobic gram-positive bacteria found in the oral cavity. Furthermore, with the rise of antibiotic resistance, bacteriophage therapy has recently been on the scientific radar. Phage therapy presents a novel approach to treat patients with dental caries, eliminating *S. Mutans* to prevent the possibility of a chronic infection. This study aims to characterize and isolate a novel bacteriophage for the purpose of targeting *S. Mutans* as well as investigate the metabolism of *S. Mutans* in various growth mediums. Our investigation involved using more than ten samples of canine saliva to isolate a bacteriophage for *S. Mutans* strain ATCC 25175 through visible plaques, in addition to recording *S. Mutans*' growth curve following supplementation with different media.

Introduction

1. Overview of dental caries and etiology and epidemiology

Dental caries remain a global concern. Poor dental health can contribute to burdens on a person's emotional and physical well-being. The most common symptom of dental diseases is pain, which is typically the stimulus to seek treatment. The etiology of dental caries can be characterized by multiple factors, including diet, acid producing bacteria, and susceptible teeth. The epidemiology of dental caries is categorized by the DMF index, which represents the total number of decayed, missing, and filled teeth. Incidence rates showing the number of DMF permanent teeth in children, adolescents, and adults show that dental caries constitutes an overall higher rate in males across the ages of 6-74. Furthermore, the connection between age and prevalence is of interest. Children ages 6-11 have the highest rates of tooth decay among these age groups. Within the 6-11 age group, the prevalence of tooth decay increases with age as the 11 year old group has more than 75% of subjects experiencing dental decay. By understanding the etiology and epidemiology of dental caries, preventative measures can be taken to address this global concern.

The current prevention and treatment of dental caries is viable, however sometimes ineffective and invasive. The treatment of cavities almost always results in removal and reconstruction, which is expensive and complex as the disease progresses. If an early diagnosis of the disease is not taken, the treatment becomes difficult as the decay can advance further into the teeth's enamel, damaging nervous tissue. Additionally, current prevention of the disease is very possible, however, unlikely for it to be implemented in the age groups which receive the

highest rates of tooth decay. Current prevention methods contain fluoride treatments, plaque removal, adhesive treatments, and diet modification. Despite the constant emphasis put on the importance of maintaining dental hygiene, children typically do not listen. Therefore, better preventative strategies should be implemented to avoid the detrimental impact of dental diseases.

Streptococcus Mutans is one of many oral bacteria which are responsible for the formation of dental plaque causing tooth decay. *S. Mutans* is a lactic acid anaerobic gram-positive bacterium classified into four major serological groups, c, e, f, and k often found in the oral cavity. Exclusively relying on glycolysis for energy production, *S. Mutans* are able to metabolize a large amount of carbohydrates. Its ability to metabolize a variety of carbohydrates and its production of organic acids allows it to dissolve tooth enamel and makes it the perfect candidate for causing caries. *S. Mutans* dissolve tooth enamel due to its unique carbohydrate metabolism. *S. Mutans* have the ability to metabolize a wide variety of carbohydrates. The metabolism of carbohydrates by *S. Mutans* comes from extracellular enzymes glycosyltransferases (GTFs) and fructosyltransferases (FTFs) enzymes which catalyze the formation of glucans and fructans that play a role in the binding of the extracellular matrix needed to form a biofilm.

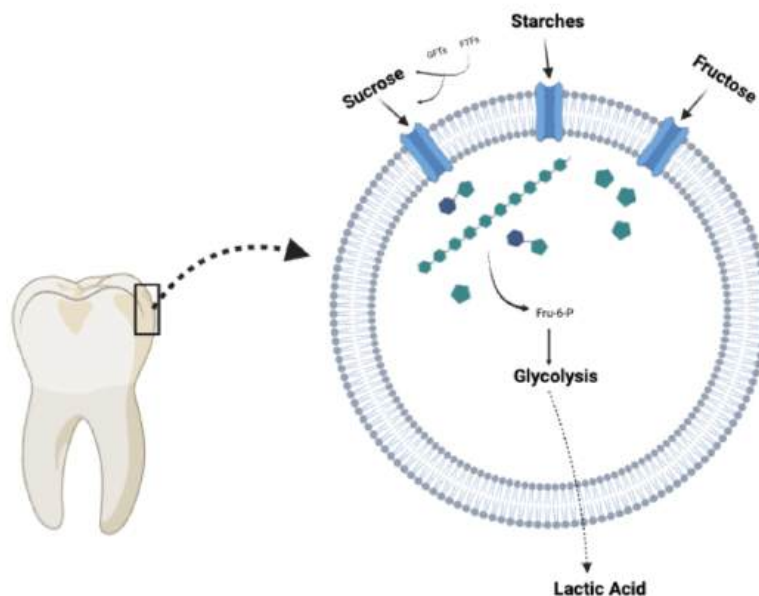


Figure (1): *S. Mutans* metabolize carbohydrates with the help of GTF and FTF enzymes that speed up the catabolism of carbohydrates into Fru-6-P which goes through glycolysis to produce the bioproduct of lactic acid.

2. Importance of bacteriophage therapy

Bacteriophages—or “phages” as they are more commonly known—are viruses that only replicate in bacteria. Recently, bacteriophages have been of scientific interest due to the rising problem of antibiotic resistance. Bacteriophages contain a nucleic acid genome in a shell of phage-encoded proteins. Most phages contain a head, leg, and tail. Phages lyse bacteria through

bacterial transduction: the mechanism of a bacteriophage in which it integrates phage DNA into encapsulated bacterial DNA in a host bacterial cell for the replication and reproduction of more phages.

Bacteriophages carry clinical significance that can be used for treating infections and disease. Firstly, they provide an alternative to antibiotics in preventative care. It is unlikely that bacteria become resistant to phages as they evolve simultaneously to combat each other. Secondly, bacteriophages can be used to detect the presence of their host in a complex environmental sample. Additionally, phages can distinguish strains of bacteria by measuring the susceptibility and resistance pattern of each phage type through panels. The clinical significance and implications of bacteriophages can be used to isolate novel bacteriophages for preventative therapy.

2.1. Role of bacteriophages in the oral microbiome

Phages are predominant in the oral virobiome, however, their role in oral health requires more investigation.⁵ The use of bacteriophages in dental health is effective due to the mouth's bacterial diversity. However, past researches have thought that the use of bacteriophages in dental health was limited. Recent studies have found that bacteriophages are more virulent to bacteria in human cells rather than to bacterial cultures which may lead to further clinical research on phages. Using bacteriophages in the oral microbiome may be particularly effective due to phage's ability to clear biofilm. Bacteriophages actively penetrate biofilms by lysing one bacterial layer at a time. Altering the bacteria which provide the biofilm can render the progression of oral infections, and due to the high prevalence of oral biofilms, phage therapy presents as a method to deal with multidrug-resistant bacterial plaques in the mouth.

Methods

3. Measurement of the growth curve of *S. Mutans*

LB-broth was made by suspending 38.2g of agar powder in one liter of DH₂O, spun and heated until dissolved, and autoclaved accordingly. Todd Hewitt broth was made following the same procedure however, with a -1g adjustment of agar. *S. Mutans* strain ATCC 25175 was inoculated and cultured using a KwikStik pellet in one 50mL centrifuge tube with LB-broth and another with TH-broth supplemented with 1g sucrose. The culture was put through the UV spectrophotometer under the 600-nm wavelength over a 14 hour time period. Data was collected accordingly.

3.1. Collection of saliva sample

A small dental sponge stick was taken and swabbed in a canine's mouth until it was coated with saliva. The sponge stick was cut and pushed through a syringe to extract the saliva sample. The sample was centrifuged at 13k RPM for 10 minutes or until there was a visible bacterial pellet. The supernatant was filtered through a 0.45 micron filter and stored at 4°C.

3.2. Double layer plaque assay Mitis-Salavarius Agar

MSA was made by suspending 18g of MSA powder with 200 mL of DH₂O. Soft MSA was made by suspending 8g of MSA powder with 200 mL of distilled water (50% of the agar concentration of hard MSA). The two mediums were heated until dissolved then autoclaved overnight. The hard agar was poured into 6, 100 mm x 15 mm plates. The soft agar was kept for future melting.

S. Mutans was cultured by using a KwikStik pellet to inoculate a 50 mL centrifuge tube of TH broth supplemented with 1g sucrose and incubated overnight for growth. The TH-broth was made by mixing 37g of TH-broth powder with 1L DH₂O, heated, and autoclaved for dissolution of the powder.

To prepare for the double layer plaque assay, 2 bacterial samples to 1 bacteriophage environmental sample must be prepared, and to properly get countable plaques, the bacteriophage environmental sample must be diluted. Serial dilution was performed as the sample was diluted into 5 samples with a 1:1 ratio of saliva and 7.8 pH phage buffer. 100 µl of saliva sample was serially diluted into 5 samples of 100 µl phage buffer.

For the double layer plaque assay, the soft agar was heated up to 55°C. 200 µl of bacterial sample to 100 µl of saliva sample with a phage buffer was poured on the hard agar. Approximately 3 mL of soft agar was then poured onto the hard agar on top of the bacterial and saliva sample. The plate was swirled to create a bacterial lawn and incubated at 37°C overnight to check for plaques. To pick a plaque, a pipette was taken and lightly stabbed in the agar for collection. The plaque was stored at 4°C in 100 µl of phage buffer.

The second double layer agar assay followed a similar procedure however with different agar. The procedure to perform the double layer agar assay was repeated with an addition of two plates plated from the past plaque. The rest of the plates were plated with the supernatant of a saliva sample which was centrifuged and diluted in buffer.

Results and Analysis

4. Measurement of *S. Mutans* growth curve

Extended Lag Phase and Exponential Growth Curve of *S. Mutans* Strain ATCC 25175 in 2% Sucrose TH-Broth in OD

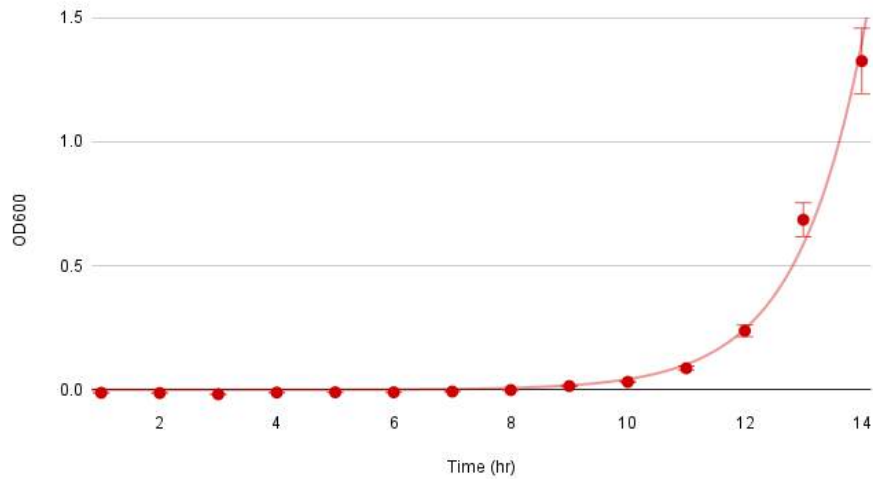
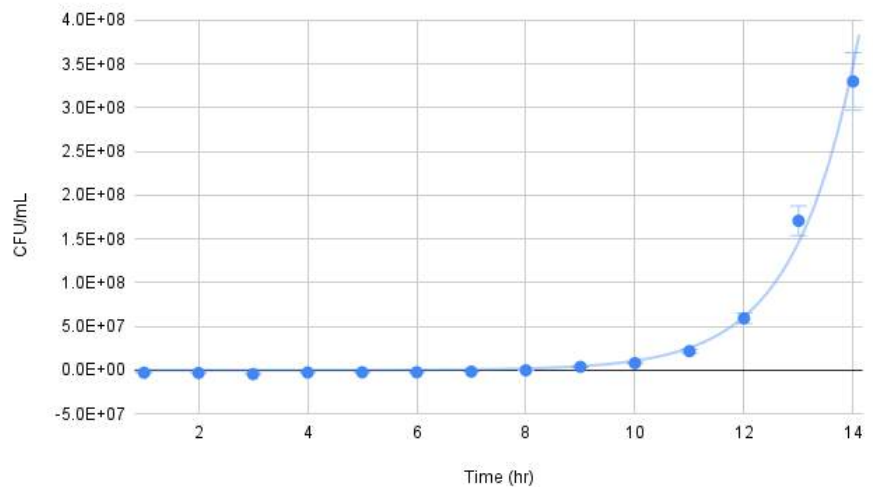


Figure (2) a: Growth curve of *S. Mutans* in a 2% sucrose TH-Broth read in OD showing the extended lag phase of *S. Mutans* in adjustment and adaptation to environmental conditions of growth media.

Extended Lag Phase and Exponential Growth Curve of *S. Mutans* Strain ATCC 25175 in 2% Sucrose TH-Broth in CFU/mL



b: Growth curve of *S. Mutans* in 2% sucrose TH-Broth showing the amount of colony forming units of *S. Mutans* throughout a 14 hour time period.

Following a 14 hour incubation of *S. Mutans* in TH-broth with 2% Sucrose supplementation, it is apparent that *S. Mutans* carry through an extended lag phase. In Figure 2 (a), *S. Mutans* do not reach exponential growth until around 13 hours, and only show growth

starting at approximately 8 hours. Figure 2 (b) is read in CFU/mL suggesting that only after 9 hours of incubation, *S. Mutans* a 1mL culture will contain enough bacteria to form small colonies.

This data suggests that *S. Mutans* have an extended lag phase in which the bacteria takes a particular long period of time to adapt to its conditions. Despite the supplementation of 2% sucrose into the media, the bacteria still fails to grow before the 8 hour incubation period. This information can be used to appropriately estimate the time it takes for *S. Mutans* to form an oral biofilm, as the exponential growth of a bacteria is suggested in the formation of a biofilm along with other factors such as the availability of resources. Finally, this data can be used in the isolation of a bacteriophage along with *S. Mutans*. In the double agar plaque assay, bacteria must be confluent enough to form sufficient colonies. Finding the growth curve of *S. Mutans* can be used to timely estimate the plate culturing of *S. Mutans* at its healthiest point; which occurs at its logarithmic phase between 12-14 hours of culture, suggesting that an overnight incubation of *S. Mutans* in broth would be sufficient for the double layer plaque assay.

4.1. Isolation of *S. Mutans* bacteriophage

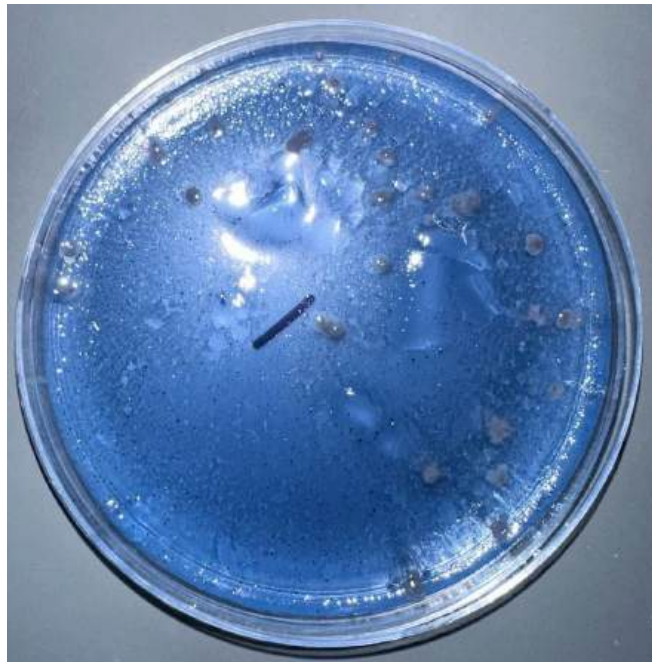


Figure (3) a: First serially diluted plate of *S. Mutans* by double layer plaque assay with saliva sample.



b: Serially diluted plate of *S. Mutans* by double layer plaque assay with phage lysate.

The colonies in Figure 3 (a) suggest that there is a bacteriophage present, however, it is unclear whether the phage has a host range for *S. Mutans*. Morphologically, there are plaques on the plate, however it is possible that the phage could be propagated from other bacteria that slipped through the centrifugation and filtration process of the saliva sample. This was suggestive in the next plates that follow the series dilution as similar colonies were formed. The plaques were picked and stored in phage buffer. Figure 3 (b) shows the bacteriophage plaque from the previous isolation in MSA in a double layer agar assay propagated with *S. Mutans*. There are no morphological plaques present, therefore *S. Mutans* are not within the host range of the bacteriophage. If the bacteriophage for another bacteria can be found in the oral microbiome, a phage for *S. Mutans* is possible. However, a previous study showed that more than 200 samples of saliva were isolated before a strain of an *S. Mutans* phage was isolated. Bacteria other than *S. Mutans* were found to have a phage in the canine saliva, therefore suggesting that the probability of an *S. Mutans* phage is substantially lower than other phages present in the oral virobiome. Future directions can be made from the isolations.

Conclusion

5. Future directions

Bacteriophage therapy in the oral microbiome can be further studied through a wider range of bacteria other than *S. Mutans*. Although the main focus of the investigation will be to isolate a bacteriophage for *S. Mutans* other strains of bacteria shall be tested as they also contribute to the formation of biofilm. Additionally, oral phage therapy can be further studied through the prevention of biofilm formation as biofilms are the main causation of tooth decay.

Furthermore, bacterial infections in the mouth other than cavities can be studied as they also present a global health concern. Despite the challenges of phage isolation, the optimization of bacteriophage therapy for oral infection is promising. This area of exploration will reveal more about the role of viruses in the oral microbiome as well as develop preventative therapies other than antibiotics to treat oral disease.

Acknowledgements

I would like to greatly thank Dr. Abascal for guiding me through this project. It would not have been possible without her patience, invaluable guidance, and expertise. I am grateful for her dedication to advising me throughout this work.

Works Cited

- Institute of Medicine. Public Policy Options for Better Dental Health: Report of a Study; The National Academies Press, **1980**. <https://doi.org/10.17226/9921>.
- Lemos, J. A.; Palmer, S. R.; Zeng, L.; Wen, Z. T.; Kajfasz, J. K.; Freires, I. A.; Abranches, J.; Brady, L. J; The Biology of Streptococcus mutans. Microbiology spectrum, **2019**, 7(1), 0-51. [10.1128/microbiolspec.GPP3-0051-2018](https://doi.org/10.1128/microbiolspec.GPP3-0051-2018) (accessed May 20, 2024).
- Kasman LM; Porter LD. Bacteriophages. **Updated 2022**. StatPearls [Online]. <https://www.ncbi.nlm.nih.gov/books/NBK493185/> (accessed May 20, 2024)
- Banar, M.; Rokaya, D.; Azizian, R.; Khurshid, Z.; & Banakar, M. Oral bacteriophages: metagenomic clues to interpret microbiomes. Microbiology, **2024**, 12, 0-23. <https://doi.org/10.7717/peerj.16947> (accessed May 20, 2024)
- Ben-Zaken, H.; Kraitman, R.; Copenhagen-Glazer, S.; Khalifa, L.; Alkalay-Oren, S.; Gelman, D.; Ben-Gal, G.; Beyth, N.; Hazan, R. Isolation and Characterization of Streptococcus mutans Phage as a Possible Treatment Agent for Caries. Viruses. **2021**, 13, 825-844.

What Is the Intent of the New Mindreading Neurotech, And How Would It be Used?

By Jahnavi Ponnolu

Introduction

In movies, we see people with superpowers which include mindreading. What if people could read minds in real life? A new neurotechnology called Brain-Computer Interface (BCI) could make this possible. This paper will also use the Kantian framework to show that BCI is a violation of a person's bodily autonomy.

BCI. What is it?

By tracking reactions to brain activity and deciphering thought patterns, implanted Brain-Computer Interface (BCI) technologies like brain fingerprinting and multi-voxel pattern analysis can identify lies and determine someone's intentions. Determining the methods for collecting, storing, and presenting brain data as evidence in court presents additional challenges. Even with all the technology's exciting potential, more thorough regulation and scrutiny are required before it can be properly included in criminal procedures. While BCI shows promise, careful analysis, standards, and more testing are necessary for its appropriate, accurate, and equitable use in high-stakes legal contexts.

Multiple distinct types of BCI technology are currently in development. These include fMRIs, invasive/semi-invasive/non-invasive BCI, brain fingerprinting, etc. fMRI stands for functional magnetic resonance imaging. This device measures brain activity by detecting variations in blood flow. This approach is based on the assumption that cerebral blood flow and neuronal activity are linked. Invasive BCI are electrodes that are surgically implanted into the brain under the scalp for communicating brain signals. The main advantage of invasive BCI is that it provides a more accurate reading than semi-invasive or non-invasive BCI. However, since it has to be surgically implanted, there are side effects from the surgery which include scarred tissue to interfere with brain signals and the body could also reject the electrodes and cause more complications. Semi-invasive BCI also requires surgical implantation. These devices are implanted inside the skull but outside the brain which produces better resolution signals than non-invasive BCI. Non-invasive BCI does not require surgical implantation. Most non-invasive BCIs depend on the electrodes, which are carefully placed in specific places on the scalp, to record brain activity. Some examples of non-invasive BCI are electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), magneto-encephalography (MEG), near-infrared spectroscopy (NIRS), and Functional Transcranial Doppler Sonography (FTDS). Using a non-invasive BCI would include wearing a device with electrical sensors that serve as two-way communication channels between a patient's brain and a machine. While this method is safer, it won't be as accurate as invasive BCI since it is not directly connected to the brain tissue. Brain fingerprinting is a lie detection system used in some criminal cases that determines whether information is stored in the subject's cognitive memory. The method is to show words, phrases, or images that highlight key details about a crime on a computer screen in succession with

unrelated stimuli to see if the suspect can identify the crime-related objects. Brain fingerprinting uses sensors that are

fMRI	Invasive	Semi-invasive	Non-invasive	Brain fingerprinting
No surgery required	Surgery required	Surgery required	No surgery required	No surgery required
Tracks blood flow to different parts of the brain.	Provides a more accurate reading since the chip is placed directly in the brain.	Safer than Invasive BCI, however, still requires surgery, but has a lower risk of complications.	Suffers from weak brain signals (poor signal-to-noise ratio). Less accuracy with the price of being completely safe.	Determines whether information is stored in the subject's cognitive memory. Uses sensors that are placed on the subject's scalp.

BCI in the Courtroom

The main applications for BCI in criminal law would be in lie detection, intent assessment and to assess criminal recidivism, the likelihood that a convicted criminal will reoffend. In this section, I will outline ways in which such technologies are already being applied as well as the case law supporting or ruling out the application of similar technologies. This will help to assess the legal barriers to implementation.

Brain fingerprinting has already been applied in two instances—one in the US and one in India—research indicates that more testing of the method is still necessary before it can be applied more widely. Even in the US case, the court did not rely on the test results; rather, it only referenced them.

The US case in which brain fingerprinting was used is with Terry Harrington. Terry Harrington was incarcerated for murder for more than half of his life. Dr. Lawrence Farwell performed brain fingerprinting tests to demonstrate, twenty-three years after Harrington's conviction, that the record kept in his brain does not fit the murder scene and does match his alibi, with a 99.9% statistical confidence level. According to the tests, he does not retain important elements of the crime in his brain. According to the judge, brain fingerprinting was acceptable in court as scientific evidence. The Iowa Supreme Court overturned his murder conviction and mandated a new trial on February 26, 2003. The State of Iowa decided not to retry Mr. Harrington, and he was released from prison in October 2003.

The primary example of a similar form of technology that is currently not admissible as evidence in court is the polygraph or lie detector test. Given the similarity, existing precedent for this technology bears relevance for the implementation of BCI.

Polygraph test results are inadmissible in court they are not considered reliable enough to use as evidence. It is well documented that certain people can pass the test while lying and

others, who are telling the truth, can fail the test. Polygraph test results may appear to be accurate all the time, however, this is not the case. Polygraph tests measure physiological changes in the cardiovascular system such as heart rate and breathing. This means that the detector can also sense when the subject is scared or nervous, but it is not able to differentiate the two. The subject could be anxious while answering truthfully but the polygraph will mistake the anxiety as lying. One specific issue is that the association between dishonesty and subjects' physiological reactions in the polygraph has not been distinguished from placebo-like effects, which are the subjects' beliefs about the procedure's effectiveness. People who think they can be caught and the test works could confess or get extremely nervous when questioned, which is one reason polygraphs appear to be accurate. We might name the lie detector a fear detector if this point of view is accurate. Not only does the court not trust polygraph tests, but psychologists such as Leonard Saxe argue that the idea of testing someone's truthfulness by monitoring psychophysiological changes is more of a myth than reality. The American Psychological Association wrote "Evidence indicates that strategies used to "beat" polygraph examinations, so-called countermeasures, may be effective. Countermeasures include simple physical movements, psychological interventions (e.g., manipulating subjects' beliefs about the test), and the use of pharmacological agents that alter arousal patterns." This shows that polygraph tests aren't entirely reliable.

BCI is also non-admissible in court as it violates the Fifth Amendment and bodily autonomy. This brain-implanted computer interface will be able to read someone's mind, but once it is in the brain, the only way to "turn it off" is to remove it. This means that if someone in the courtroom has the BCI implanted, it will violate their Fifth Amendment right which states that a witness can "plead the fifth" and protect themselves against self-incrimination. This is true for all BCIs, both invasive and non-invasive. Once they are connected they have full access to your thoughts; there is no limit.

In "Reading the Mind: Neuroimaging Technology and its possible use in Court" Cumming (2020) writes about how BCI can be used in court to prove intention. The article states that criminal case-solving has already benefited significantly from technology. Since its introduction into the UK in the 1980s by Sir Alec Jeffreys, forensic DNA analysis has grown to be an essential component of criminal investigations and has influenced their methodology. While many crimes have been solved as a result of this, a crucial component is left out: the accused's intention. A criminal case consists of two main components: *actus reus* and *mens rea*, or the act and the intention. Establishing the intention can be difficult, particularly when the defendant has oblique intent, which occurs when they take a course of action that results in the desired outcome but they are aware that it will also have unintended consequences. This article uses the example of placing a bomb in a car to detonate and murder the occupants. The criminal charge itself may vary if the defendant set a small bomb with the aim of simply causing damage to the car and was shocked to see that it killed the person inside.

“Use of Neurotechnologies and Neuroscience in Legal Settings: Case Studies” by The National Library of Medicine talks about neurotech within the legal system. It uses quotes from people like Joshua Sanes and Joshua Buckholtz. Joshua Sanes writes that in the future, technologies might be able to predict someone's dangerousness or likelihood of recidivating, evaluate their volition and intent, establish their competence to stand trial, identify biological factors that could mitigate criminal behavior, separate false memories from real ones, recover lost memories, or discern between genuine and fake memories that have been recovered, according to Sanes. He continued, "They also present the opportunity to maximize treatment and lower recidivism." Sanes foresaw swift and startling technological advancements that would make things more interesting rather than necessarily more dependable. However, according to Joshua Buckholtz, there isn't much overlap between what the law believes neuroscience should do or could do better than what is currently being done, what people—including some scientists—have claimed neuroscience can do, and what neuroscience can actually do, either now or in the future. Furthermore, he stated that no systematic assessment of the overlap has been carried out, therefore no framework for harmonizing legislative objectives, societal norms, and neuroscientific practices has been agreed upon. As a result, nobody is certain what belongs in that intersection.

Buckholtz listed three areas of interaction between law and neuroscience to develop such a framework: “disclosing mental states, assessing self-control ability, and forecasting future behavior.” Brain imaging has potential in these areas, but Buckholtz advised setting reasonable expectations. According to Buckholtz, mental states that might be clarified by brain imaging methods like fMRI could allow for degree of intent identification, bias demonstration in a witness or juror, suffering quantification, or lie detection. According to Buckholtz, there are three components to the capacity for self-control at the nexus of law and neuroscience: (1) assessing maturity, (2) identifying impairments, and (3) verifying the existence of illness states that are important to the law. There hasn't been a logical mapping between the legal and scientific conceptions of self-control, he claimed, because neuroscientific evidence in this field usually aims to reduce culpability by inserting biology as a mitigating element.

To clarify the differences between the legal and scientific frameworks, Buckholtz provided a fictitious case study of a man accused of a violent crime who did not score well on measures of purposeful control, reaction inhibition, action cancellation, or impulsive choice. According to Buckholtz, the language of the law is incompatible with neuroscience measurements, thus it is now difficult to conclude that a man's performance on these tests indicates a mental defect that produced a significant incapacity for a man to conform his conduct to the requirements of the law. Though concepts like irresistible urge or volitional capacity have no value to cognitive neuroscientists, according to Buckholtz, phrases like "action cancellation" and "response inhibition," which are commonly used in cognitive neuroscience, have no bearing on those who formulate laws. He noted that for policymakers to benefit from neuroscientific knowledge, they must carefully consider whether any of the self-control domains that neuroscientists can access have an impact on legal responsibility.

The Ethics Behind it All

Since the BCI is a violation of the Fifth Amendment, is or was it ever considered ethical? Not only does it violate an individual's Fifth Amendment right, but it is also a violation of bodily autonomy. According to Kant's Autonomy Formula, "the Idea of the will of every rational being as *a will that legislates universal law*." Here, however, we emphasize that we are global law providers, not just followers. Self-interest and other natural, immoral motivations could lead to a rational will that is only constrained by universal laws to behave in a certain way. However, giving up such conditional motives—motivations that rational agents like us might or might not have—is necessary to be a legislator of universal rules. In light of the rational will's autonomy and its place as the source of the universal rules that bind it, this formulation requires us to conduct ourselves in a way that is consistent with these principles. Now we ask the question, according to Kant's Autonomy Formula, is BCI ethical? While its intent is simply to try to improve our communication or to help someone in a serious situation, it could always become more. With this kind of technology, people can do all kinds of things, and when the chip is implanted in your brain, all your privacy goes out the window. Your mind will become a glass box with people always looking into it. The chip won't be able to filter some things out and just show what is needed, *everything* will be shown. BCI, in the lens of Kant, is not ethical because it violates an individual's right to bodily autonomy.

Conclusion

In conclusion, all kinds of BCI have the potential to revolutionize the way we live and our legal system, it's just a matter of developmental, legal, and ethical concerns. We don't know exactly how accurate invasive BCI is, so it won't be a reliable enough test, similar to the polygraph test, in court. By using BCI in these high-stakes situations, the defendant is at risk of being either falsely accused or of getting away with a crime they did commit.

Works Cited

- Becher, Brooke, et al. "Brain-Computer Interfaces (BCI), Explained." Built In, <https://builtin.com/hardware/brain-computer-interface-bci>. Accessed 28 July 2024.
- "Brain Computer Interface (BCI) System Overview & Applications." How To Electronics, <https://how2electronics.com/brain-computer-interface-system-overview-applications/>. Accessed 28 July 2024.
- Cumming, Alex. "Reading the Mind: Neuroimaging technology and its possible use in court." SCL Student Bytes, 14 September 2020, <https://bytes.scl.org/reading-the-mind-neuroimaging-technology-and-its-possible-use-in-court/>. Accessed 28 July 2024.
- "Ethical Implications of Invasive BCI in Law Enforcement Surveillance." Medium, 23 February 2024, <https://medium.com/@innovirtuoso/ethical-implications-of-invasive-bci-in-law-enforcement-surveillance-609d7a1c77c0>. Accessed 28 July 2024.
- "Harrington Case." Farwell Brain Fingerprinting, <https://farwellbrainfingerprinting.com/harrington-case/>. Accessed 30 August 2024.
- Johnson, Robert, and Adam Cureton. "Kant's Moral Philosophy (Stanford Encyclopedia of Philosophy)." Stanford Encyclopedia of Philosophy, 23 February 2004, <https://plato.stanford.edu/entries/kant-moral/#Aut>. Accessed 28 July 2024.
- "Use of Neurotechnologies and Neuroscience in Legal Settings: Case Studies." National Library of Medicine, <https://www.ncbi.nlm.nih.gov/books/NBK519347/#:~:text=Mental%20states%20that%20might%20be,or%20detecting%20lies%2C%20said%20Buckholtz>.

Is Silliness Required for an Economy? By Vihan Agrawal

Abstract

Luxury goods or Veblen goods are goods that are not necessary to live but are deemed highly desirable in society. Some can deem spending huge amounts of money on goods like – modern art as foolish. While others emphasize that they get quality as value added. The market of luxury goods was valued at \$284B in 2023 and this is just the figure for luxury apparel wear. When I refer to the luxury goods in the paper, I refer to all the goods that we buy but are not necessary like \$1000 water bottles and many more. These goods play a vital role in the economy and so also have a huge implication on it. Although the demand for such goods is highly elastic with their demand being less in financial crisis there is always some demand for some luxury goods in the long run. The reason why people buy these goods is not just because they have money but because there is behavioral economics and psychology involved behind it. The research paper aims to find the psychology behind it and how companies exploit it and with data try to predict future trends by analyzing a scenario of a bubble burst by understanding the factors affecting the fashion sense of the people and also the consequences of getting rid of them on the economy.

The research paper essentially the end wants to evaluate whether everything having utility is important or not and does the silliness of the people might just be necessary.

Who and why do people buy these goods?

The foremost buyers of these goods are the middle-class groups which account for nearly half of all the sales. The people who can't afford these and barely afford these with months of savings are the ones who buy these the most. When people buy these goods, they feel a sense of accomplishment and they feel rich. They think that buying this will make them happy. But this tendency in itself has many economic consequences. Americans in 2024 have \$1,115 trillion¹² of credit debt. The increase in debt is a concern. This also depicts the irrationality of the consumers and their tendency to indulge in things they can't afford. As credit cards have high interest-rate a small debt for a luxury good can turn into a big debt if not paid on time. Another thing is that people buy goods looking at their neighbors, if the neighbor has bought expensive things like a new car etc., we would also want to buy the same things as a measure of success. People also buy these goods as they feel respected and good against others in society with cheaper clothes. In some cases, this makes them feel good with a stamp of social approval, an announcement of being superior to others and gives them a spot in society¹³. There is also an ostrich effect at play and so whenever someone tells them about how there is nothing much that differentiates them from the other products, they ignore this information and choose to believe that higher price is

¹² Schulz, Matt. "2024 Credit Card Debt Statistics." LendingTree, 9 Aug. 2024, www.lendingtree.com/credit-cards/study/credit-card-debt-statistics/#:~:text=Americans%20have%20an%20absolute%20mountain,credit%20card%20use%20each%20month.

¹³ Bochańczyk-Kupka, Dominika. (2019). NECESSITY AND LUXURY GOODS IN ECONOMICS. 10.32008/NORDSCI2019/B2/V2/16.

equal to higher quality. Also, as people use fashion as a personal statement about themselves in the present world where personal branding has become so important they use these extravagant goods to express themselves and create that type of persona around them. In current times with the effect of media and influencers being the highest, people after seeing them with those luxury items want the same thing because they want to be like them. How irrational it might seem. The top one percent don't buy them the most because they are already rich and everyone knows it but people of lower sections of society many seem to want to get that type of attraction of being rich. This is not just a recent thing; this has been happening since ancient times like when printing was first introduced as books were able to be mass-printed and the books got relatively cheaper with hand-written books being more expensive, so the rich to show their status brought these books and then got them painted and designed by artisans. They did this to be different from the poor and in some ways, this is similar to how in today's world people are buying stuff even in debt to be different from the poor. Does increasing awareness about it help much? The answer is no. Many people who couldn't buy it at first even after months of saving justify these goods as expensive but then as their income rises the same people justify it as worth it for quality and design or saying it's okay to treat ourselves once in a while. This proves how people even after knowing these goods are worthless later justify them showing how much the environment created around it by the brand and the society affects people's minds to such a great extent.

How do companies profit from it?

The companies feed on these exact emotions and try to create a sense of exclusivity around the brand. They make you feel special with good stores and polite and good customer service. They keep their supply limited to not make their things common so the consumers feel special and different. They leverage their higher prices to create a prestigious image. They use celebrity endorsement to make sure their products are worn by influencers and celebrities because it appeals to GenZ, the largest consumer of these goods, be. After all, you know seeing it targeted audience will also want it so they create hype around their products. They create scarcity artificially to make sure the demand remains high with limited supply.¹⁴ Then the introduction of BUY NOW, PAY LATER which lets them pay in installments making them seem more accessible for middle-class income people is another factor for the increase in demand and debt when they can't afford these things and face any problems while paying installments. Further, they launch limited editions to create further scarcity and hype, then sell them at even more higher prices. Again, people want for the exclusivity when they can buy the same thing for much cheaper. The counterfeit market for such goods is more than a trillion as of 2023 which although raises many problems with these goods potentially being harmful to the health of consumers and damaging the reputation of businesses also raises the question of how to distinguish fake from real. Proving how there is not much difference between them. Then there was the 'Palessi' stunt

¹⁴ **"The Psychology Behind Why People Buy Luxury Goods."** N.p., n.d. Web. 3 Sep. 2024
<<https://www.investopedia.com/articles/personal-finance/091115/psychology-behind-why-people-buy-luxury-goods.asp>>.

where they put cheap shoes from Payless for under \$100 in a mall with a very good showroom with luxury brands like setting and also high price tag and brought influencers and other people who wear these goods there and then asked them about the products. People thinking that it was a premium brand with only this much work started praising it for how good they are and worth it for thousands of dollars proving these big brands are just overpricing their goods and also proving the irrationality of the consumers. Also, there are many cases of these brands exploiting workers and making them cheap and then just slapping a tag of thousands of dollars. In reality, the so-called “Best quality leather” comes from countries like India, China where they buy it for cheap set up factories with only minimum wages and make the complete product under \$100, then import these goods to the USA where they put their logo on it and then slap price tag of thousands of dollars. With labor laws being quite lenient in India and high rates of unemployment, these brands are easily able to do stuff like this and profit thousands of dollars from these cheap goods.

Is the luxury goods bubble going to burst?

An overpriced asset bubble is not a new thing. A similar thing happened with the tulips¹⁵ in mid mid-seventeenth century in Holland. When the price of tulips got so great and not just any tulips it was the tulips with viruses (the viruses made them look pretty with some natural design) were the ones that were demanded the most in this tulip mania. The viruses lived in the bulb of the plant and were not necessarily going to reproduce in another plant grown from the plant with the virus.

These made them rare i.e. low supply and wanted due to their beauty i.e. high demand. Also, the virus made them fragile increasing the risk. This makes them very precious with many people investing in growing them and losing money with the virus also having other drawbacks. At, last this bubble burst with them losing their value as they couldn't be replicated anymore. Coming back to the present time, although the luxury goods of the present time can still be replicated now with overpricing there are many counterfeits too in the market which is reducing the demand for these goods with the counterfeit market going larger and larger with its valuation being estimated around a trillion. This is making luxury companies lose their money as the goods no longer remain exclusive with people trying to buy more counterfeit instead of taking more debt as they are no longer distinguishable. This still makes them show off the money that they don't have while not spending months of saving and still having a luxury good. Also, with a lot of credit det already created mounted they can't take more debt which can be counted as another factor for the decreased demand. These reasons have been reducing the sales of many luxury companies nowadays. With especially the prices of luxury watches have fallen especially since 2022. Whether this trend will continue and the bubble burst is uncertain but seeing the consumer psychology some sort of overpriced things that distinguish them from the poor will always remain as we saw above.

¹⁵ Marples, Alice. “The South Sea Bubble of 1720 - The National Archives blog.” The National Archives Blog, 18 Sept. 2020, blog.nationalarchives.gov.uk/the-south-sea-bubble-of-1720

Consumerism – The start of silliness

Consumerism¹⁶ slowly started in the 19th century. With the turn of the century, coming up with small simple tools and then industrialization. We have always consumed the necessities needed to live. But the society slowly started growing and becoming more complex. Wars and food shortages also started happening as the society grew. Then frugality and thrift started becoming a more common phenomenon in a place where even the survival ratios were not guaranteed. In the late 18th century Britain took place the start of consumerism where the attempts to promote new fashion by harnessing “the power of envy” and boost sales took place. The people with money started buying mirrors, heavy funny dresses, and wigs. The church started calling these people heretics and started opposing them not wanting people to indulge in these so-called “frivolous things”. People started wanting more and some wanted to show that they were above the other comments and were special. This fueled the economy with more aggregate demand and caused growth. Small industries started coming up and then industrialization took place. Although fueled with envy this was a major step with led to the growth of human civilization and brought us here. If it was not for this fueled demand we would not be here sitting in the A/C reading this on our laptops.

What if it never happened?

However with human society becoming more complex and coming up with many philosophers, scientists, etc. It became impossible for consumerism to not happen sooner or later amidst wars and food shortages with greed being embedded like human beings. But what if it never happened? Let’s go back to a time when although society was growing bigger there was trust and love among society. The people lived simple lives without much need. Although there was still a hierarchy in society. But then without the start of consumerism, the growth of the economy would have not kick-started and with only basic products in the market the incomes would not have started rising with population growth without much demand there would have been high unemployment and starvation lead to lot of deaths acting as a population control. The only thing distinguishing the rich from the poor is the piece of land one holds. There would not have been enough funds or even basic tools to make even something useful from the works of various visionaries and studying would just have been sciences and mathematics would just have been a past-time of very few in the world but that would have remained to just theoretical work not being able to materialize without the power of envy and funds with people only living for necessities and rich having lavish lives from what labor in the farm harvested. Even if some useful things and simple tools would have materialized there wouldn’t be enough money in the market to buy them and they would just have been used by the royal household for various purposes. Maybe even with these small useful inventions civilization would start growing with the need for power and developing of better weapons as needed and funded by royals and nobles this would have been devastating with much destruction and wars,

¹⁶ The School of Life. “HISTORY OF IDEAS - Consumerism.” YouTube, 7 Oct. 2016, www.youtube.com/watch?v=Y-Unq3R--M0.

as useful projects would have been deemed to be just a dream and others not worth spending upon by royals as there won't be any profit for them. It would have taken a thousand years to get where we are now but keep in mind that luxury goods never materialized all the time. Only possible if people were more rational consumers and even with complexity growing, they wouldn't be attracted by consumer goods. And also, without much need for money and a way to differentiate themselves.

Also, luxury goods appeal to consumers more than we think, they act as a goal for the consumers to work to achieve. They work so that they can afford these useless goods. The people are motivated to work to be more productive earn more and hence be able to afford luxuries. Without the need for luxuries when even kids don't cry about buying them these useless things- Why would anyone work so hard? Why would they think so much and try to be more creative in coming up with new ideas? There would be nothing pulling people to work harder hence There won't be so many new ideas and rapid development anyways other than the problems mentioned above.

Usefulness and Uselessness¹⁷

So far, we have characterized silliness as the requirement of things that serve no purpose and which are therefore useless, such as the desire to have all the products made. But how do we draw the line between usefulness and uselessness? It is paramount to comprehend well enough usefulness and its deeper philosophical dimension usefulness to appreciate the reason for uselessness and therefore silliness in the development or economy in perspective of material somewhat agreeable to what we have so far discussed which is the provision of only the seemingly useless goods and knowledge production economy that seems to be economically unproductive. This notion of uselessness and silliness may be said to have been misapprehended and this is the need for a broader definition of usefulness and uselessness and their relevance.

Usefulness

The term usefulness cannot be addressed with a headache in a society where most societies do not subscribe to what is perceived as societal norms. Democracy encompasses a fairly contested topic where issues of judiciousness and rationale cut across more than society's political corner. More particularly, how functionalities are understood, has become an affront in modern societies, works in question especially their applications in philosophical and humanitarian domains. From that very day, usefulness dominated over subjective evaluation of the value of any activity including experiences and relationships. The consequences of this emphasis on utility and efficiency are that processes or activities that cannot seem to generate objective results are often branded as wasteful or unproductive. Simplicity notwithstanding, explaining usefulness is less of an easy task. At the most basic level, a useful object is defined as one that can be of use to someone and for what purpose. This definition brings out the fact that usefulness is relational – there are no 'useful' things in and of

¹⁷ Šliuburys, Vytautas. At the margins of productivity: philosophy of uselessness. Diss.

themselves, there are only useful things that are aimed at some end about some user. The term usefulness, however, implies more abundant meanings than the presence of a predetermined use. Identifying an object or an idea as useful does not always mean that it has been used as intended. It is worth noting that some theorists are eager to stretch the definition of utility to mean any kind of benefit that can result from interaction with something. This broader outlook would incorporate theoretical or philosophic work and creative work as well and consequently tend to justify even philosophy and arts within the utilitarianism paradigm. Yet this methodology in performing arts perpetually disturbs the clarity of the value of the concept of usefulness to the extent that it loses its purposefulness.

In the age of productivity mania, people equate being useful to being happy, or to one's worth. One should seek out 'useful' events and connections, discounting those opportunities and activities which, though may be pleasurable, may not bring any 'useful' results. Such trends may mean a contraction of the human condition and a contempt for things that remain unproductive or do not have flourishing success. The usefulness of a particular ideology seems entrenched in modern society and presents difficulties for fields such as philosophy and the humanities. Such disciplines, which, more often than not, do not offer immediate useful or marketable results are becoming more and more irrelevant in a society that prioritizes utilitarianism even as the Veblen goods are viewed as still useful due to their value. Others think of it simple misplaced effort to try and bring such fields into the stereotypical box of 'useful' since doing so only serves to erode their very essence.

Some thinkers, rather than trying to incorporate philosophy and other "useless" disciplines in terms of usefulness, would suggest accepting and even reveling in their uselessness. This way of thinking has the implications that there are things that are useful that fall outside utilitarianism and help people to achieve more meaningful participation in aspects of life that help them to experience a more sovereign self, more than just success in the context of consumption such as feeling accomplished when one has saved up money to purchase their first designer clothes like Gucci or LV, especially for the middle class. It follows, therefore, that the controversy of usefulness and its relevance in society generates a discourse that agitates for more fundamental human concerns which include fundamental beliefs about people, what it means to be fulfilled, and how human beings deal with the practical versus wider requirements of the full human being. While usefulness is not an issue, there is a zone that it should encircle. Concentration on utility alone tends to be detrimental to the cultural and intellectual aspects hence deafening other channels of innovation, thinking, and acquiring knowledge for knowledge's sake.

As a consequence of constant emphasis on result-oriented attitudes, measurable outcomes have become more common as well in modern society. This tendency seems to be particularly strong within economic and educational systems in which efforts lacking in tangible or clear economic returns are most of the time viewed as hindering development and improvement. People in modern society are persistent in the attainment of objective numerical goals. One common metric of success in many industries is indexed around productivity levels, profitability

levels, and levels of efficiency. This measurable outcomes concept is exacerbated by the need for making more economic growth and deploying resources optimally. In such a setting, what is not favorable in generating tangible or visible economic returns is typically dismissed as wasteful or unproductive.

Economic Efficiency and Opportunity Cost

The opportunity cost as embedded in economic theory provides a rationale for measurable outcomes. Opportunity cost is when a certain choice is made, and then the loss of the value of that particular alternative is thrown. It helps by concentrating on cases with a discernible economic payoff thus more resources including time, labor and capital will be available for effort and growth. In these terms, economic efficiency refers to a situation where all available resources are optimally applied in the production of wealth and improve living standards.

The pros and cons of post-modernism are art and design utilitarianism. Nevertheless, this utilitarian approach has significant limitations. One major disadvantage is that activities that do not call for a particular type of economic activity are pushed out of the equation, it focuses on the immediate benefits. Such fields are philosophy, humanities, and the arts. These areas, therefore, are regarded as "ineffective" as they do not bring in any monetary value. It is quite obvious, that they are yielding less money, popularity, and word of mouth in the type of society that is cash-driven. Not many appreciate the modern art that the rich or the middle class buys for decorative purposes or for bragging about the excessive price tag. It can be accepted as useful by those who possess the understanding of the art's value and the knowledge embodied in the art creation though others don't and thus often regard it as an item for ostentation.

Work of this type may not have a direct financial impact on any strategic objective: Structured Work / Non-Quantifiable Work – Broader Value of Experiences, Activities. Even if in these cases, their contribution may be less visible than monetary, these attendant activities labeled as 'useless' actually create positive value. For instance, in philosophy and the humanities critical thinking, moral judgment, and understanding one's culture are developed. These skills are vital in addressing 21st-century education needs and can foster creativity, integration into society, and personal satisfaction. In the same way, the arts have a positive impact on our lives in terms of offering aesthetic enjoyment as well as stimulating creativity, and providing fresh insights into humanity.

The Pitfalls of Focusing Too Much on Measurable Objectives

One of the drawbacks of overemphasizing measurable outcomes is that it embraces several negative consequences. Firstly, it may be detrimental to both creativity and innovation. It is no secret that some of the most stunning ideas and inventions have been realized from pursuits that were usually regarded as 'useless' or impractical. Thus, by undermining such activities, society runs the risk of not being able to fully explore its growth potential.

Secondly, the quest for tangible results can promote a very simplistic and pragmatic view of education and human progress. It is possible that education systems when it comes to

professional training, focus on metrics in terms of how many students graduate with a degree rather than instruction about critical concerns of education. This may result in a population that possesses the necessary technical skills for work, but the population lacks skills for creative thinking, complex problems, and ethical reasoning.

Third and also, on the opposite wing, the pressure of proving oneself in terms of achievements, thanks to an ever-increasing emphasis on measurable outcomes, may lead to hyper-productiveness and consequently a culture of burnout. Working in a society whose highest value is productive work can make people feel that they must work, achieve, and create all the time otherwise it is stress, anxiety, or hovering somewhere around mediocrity. Seeking to attain such a high level of economic efficiency unfortunately may be detrimental to the well-being of individuals and the functioning of social systems. This argument for a redefinition of usefulness seeks to protect those disciplines and practices that are often left aside in an ever more pragmatic society. Supporters of this position assert that spheres like philosophy, although not much useful in economic output in a direct way add much intangible value which is very important for the advancement of the society and the growth of the individuals as well. One such importance of philosophy is critical thinking development which justifies how the philosophy is practicable, generally referred to as practical philosophy. Philosophy promotes such capacities as willingness to challenge unjustified claims, logical evaluation of opposing views, and taking into account other positions – all of which are useful in many business and personal situations. Such abilities could enhance such processes as a confrontation of at-hand issues, making decisions, and creation in various areas. Also, many philosophical questions deal with the issues of ethics, morality, and the human condition. Such proponents claim that analysis of such issues helps build more responsible, kind, and moral people and culture. The capacity to respond to such questions has never been as important as it is today, given the prevalence of complexities and interdependence in the world community.

Culture, including philosophy as well as other aspects of the humanities, conservation, and management of culture turns out to be significant as well. Differences between human beings and their societies, morals, and history help in developing policies, progressive societies, and cultural growth. From this angle, this ‘usefulness’ is not limited to the more useful features of textbook philosophy. And yes, theirs is the point as far as the usefulness of philosophical thought is concerned, the advance of the sciences and technologies included. Philosophical understanding and philosophical principles have made great contributions to many scientific theories and encouraged the technologies of the day and the coming days as well. In this way, it could open new domains in other disciplines of inquiring by encouraging fantasies as well as imaginative abstractions. Nevertheless, this type of reframing usefulness that has some critics to this approach should be taken with caution for more than a few reasons. They also claim that just as individuals try to explain philosophy and other activities in terms of practical utility such people should not be in philosophy or such activities. It is the central topic that seems to arouse and capture their most lively interest which is accountable for why they center only on its elucidative

role. Institutions however should be separate from philosophical inquiry as the reason for which knowledge and understanding are sought ought not to be for pragmatic purposes only.

Secondly, there is the possibility that while attempting to categorize philosophy into a box of usefulness, we end up violating its very character and character. Some philosophy goes against the values and systems that may not be too efficient all the time. Attempting to do so may prove detrimental to the very essence of the philosophy and its innovative character. Furthermore, the criticism of this defect is that, without intention, it produces the very criterion it claims to resist, which is, that everything must be defended on the ground of efficiency.

In other words, if we strive to demonstrate a statement in philosophy, more probable would a presupposition be upheld that valuable objects are only those which have a practical application. In addition, there is a possibility that as we stressed on the indirect impact of philosophy, its direct impact may be neglected or even undervalued. Engaging in philosophical inquiry shall bring forth insights and ideas able to develop new ways of looking at things and doing things that focus on intermediation, yet these benefits in most cases are overlooked. Also, some defend themselves from the argument that seeks to change this prevailing attitude by presupposing that value should be conceptualized in broader terms and not in terms of its utility value alone. Rather than seeking to leverage the pragmatic aspect of the usefulness of philosophy, they propose that society must be engaged in expanding the scope of what is considered value and worth outside the aspect of utility.

Uselessness and its significance

One of the main weaknesses in the utilitarian view of people's actions is their neglect of the concept of what George Bataille calls "sovereign moments." It is within that context where Bataille's philosophy may be much more relevant – he asserts the need for such usefulness transcending the mundane existence-oriented spirit. This concept demolishes the conventional endorsing ideology that all human undertakings ought to be appraised with the economic viability feature. Central to any analysis of Bataille's philosophy is that human beings are more than just workers and calculators. He holds everybody possesses an internal urge to reach out to those activities which are more than simple practicality – pleasures, that are deep and ecstatic, erotic and even excessive. Such "sovereign moments" are free from the everyday obligations and the expectations that society may impose on the individual. For instance- the satisfaction one gets when buying and driving a brand-new expensive car when it is clear that the powerful engine is not going to be used since it will be driven in a city with low-speed limits.

Bataille posits that such pursuits or actions usually viewed as "unproductive" in a utilitarian sense – for instance, creation or appreciation of art, participation in religions, or engaging in luxurious pleasures – can help in reaching these sovereign moments. In these kinds of moments, people do not have to concern themselves with producing or creating any specific result. Rather, they can touch and taste life in all its richness, without the perpetual need to be productive or of any practical use. These sovereign moments as defined by Bataille are also related to his concept of a general economy in which every society is capable of generating more

“energy” than what is productive. From this perspective, even expenditures that would appear extravagant or redundant to the present understanding of economics are critical for psychological equilibrium as well as social equilibrium. They provide an outlet for the overabundant energy and urges that exist in a society that solely concentrates on production and accumulation.

Bataille considers these sovereign experiences as more than mere excesses and as something vital for self-actualization as well as creativity. The only way people will go beyond the ordinary way of living which is purely driven by the need of being practical and functional, is through blurring. New insights and ideas as well as creative ways of doing things that may not be found in pure reason or actionable thought may be developed. Besides, these sovereign moments of experience can also be beneficial to the economic and social issues of the society according to Bataille. Fulfilling these sovereign moments gives people the feeling that other productive activities beyond the routine are exploiting them. The very sparks of new ideas and business concepts that will be productive energizing creativity and innovations will likewise be developing the economy and social systems.

Let me highlight that the concept of the sovereign moments as associated with Bataille does not suggest a blanket prohibition of utility or production. Instead, it offers a broader perspective of humanity in which productive and non-productive activities equally matter. Within this effort, practices viewed as ‘wasteful’ or ‘non-productive’ from an economic standpoint would be constructed as both necessary and beneficial to a complete person. Supporters of Judith Butler and Costa M. Mainwaring might say, on the contrary, that an emphasis on these kinds of ‘useless’ activities can create economic inefficiency or social disorder. Such a society, they would maintain, risks becoming stale and oppressive if it does not allow its members to engage in these sovereign experiences and would ultimately be unsustainable

At this time, it makes me particularly anxious and restless because it invites me to uphold Bergson's contractor interest in a line of disciplined analysis that purely defies satisfaction for desires. That way, when we see the worth of activities that don't appear to be useful, we start to create better opportunities for employment, creativity, and sounder development practices. Most probably, these thoughts would appear to be fundamentally wrong in the context of ordinary perceptions of what productivity and progress mean. Today, due to their low ability to generate an actual product, philosophy, and the humanities are often considered useless. However, these fields are fundamental for providing one's analytical ability and cultural influence that may foster social advancement and innovation. They bring a value that cannot be measured only by current values.

Philosophy and Humanities and Their Contribution to the Development of Critical Thinking Philosophy and the humanities stand out because their primary contribution is the sharpening of thinking. They make people reconsider the basis, evaluate assertions, and appreciate various standpoints. This is key in any form of problem-solving concerning individual, corporate, and even societal issues.

Philosophy, for example, is a discipline that promotes reasoning and ethical consideration. These are useful particularities in professions that are highly decision-making oriented, for example, law, medicine, and policy-making. Addressed in historiography, sociology, literature or other sciences humanistic aspects provide people with critical and vibrant perceptions of the world and themselves. These are important for understanding the intricacies of the contemporary world and creating active and responsible citizens.

Philosophy and humanities do not only enhance critical thinking but also add to the culture of particular regions or people. They add value regarding past facts, belief systems, and the organization of society. Understanding these cultural elements plays an important role in building understanding, acceptance, and harmony within societies. One of the goals of learning about different peoples and times is the search for understanding people and society. The ascribed benefits of social sciences and humanities, and thus, education on the whole, undermined economic development, and shortfalls through a direct lens, cannot be overlooked. Information knowledge and culture are critical modifiers and stimulants of economic activities. So for risk undertaking modern innovations, citizens' skill in the use of practical knowledge created for critical thinking enhancement is an underlying life factor. Forecasting and helping to develop ways of optimizing processes are a must for scientific or technological progress, hence a requirement for policy, planning, or engineering processes. But the cruel truth is that empirical academics tend to ignore, if not settle allergic to, philosophy, let alone more practical matters or their vision of "social sciences". Rather they tend to fall into establishing as long as it is practical covers, but blindly imitating what has been done by others, abandonment of ways of too much politicization by the 'conventional politics. The humanities also help foster a comprehensive workforce. Of increasing importance to employers are soft skills, for instance, communication, cooperation, and moral judgment, which are enhanced through the study of philosophy and humanities. Such skills are a firm basis for economic productivity in leadership, cooperation in teams, and even at the customer interface.

Apart from these advantages, the importance of philosophy and the humanities is often challenged in the contemporary world, especially when efficiency and quantifiable results are demanded and idealized. Detractors assert that such pursuits are only an exercise in futility, because there are no real, practical results to be witnessed, and should thus be funded less and supported. They underestimate the impact that the humanities bear upon the individual and society by folly attending to a more efficient way to solve a problem.

Furthermore, the effort to shift the focus towards the indirect benefits of philosophy and the humanities to justify their existence can be observed as apologetic. Critics argue that this approach reinforces the narrow conception of value that it seeks to challenge. Rather than attempting to align these disciplines with a utilitarian purpose, one should appreciate their value per se and promote a different conception of value.

Philosophy and the humanities provide the analytical and cultural skills that are essential for the creativity and development of society. Although their direct economic impacts may appear minimal direct impacts, their repercussions are considerable. These disciplines develop

critically and tomorrow's spheres of creative activity, which are important in solving the niggling problems that the society is usually faced with. As a result of taking into consideration and appreciating the more extensive meaning of its importance to society, philosophies, and humanities, we can also accelerate a non-myopic and non-partial growth expansion. This justified obsession with disembodied productivity and usefulness in today's society has created an illness of hyperproduction that although helped in furthering the economy and looks good on report cards, is bad for individual bodies and indeed even for societies in the future. This essay revisits the subject of hyper-productivity by introducing a concept that has not been given due attention as it deserves, the concept of 'strategic uselessness' which has purportedly many useful sides in enhancing one's welfare and economic development.

The super-productivity paradigm is something that has crept into almost all spheres of contemporary life. Whether at work or pursuing one's personal development, there is a continuous demand to work to one's maximum capability, relate the time spent against the output, and work towards producing something of merit continuously. Usually, such a mindset is bred by society, business cultures, or even self-guide books that preach that one has always to be busy and achieve something.

Nonetheless, prioritizing productivity and value has its disadvantages. Several researchers have established that extreme working hours combined with the pressure to work optimally can push an individual to develop what is referred to as burnout which is a health condition characterized by fatigue both physical and emotional. A Gallup report in 2018 showed that almost 23% of employees are burned out very often or always, while an additional 44% say that they feel the same sometimes. This kind of burnout can be expressed through depressed levels of performance, increasing rates of absenteeism, or health problems among other things. In addition, acceptance of excessive performance demands can bring disappointment and unhappiness. Citing Byung-Chul Han in 'The Burnout Society', people would no longer have time for leisure activities which in turn would deprive them of rest and tranquility that are critical to a balanced way of life since the focus at all times will be productivity. According to Han, an "achievement society" does this because it indulges its sick members and patients in incessant and never-ending self-improvement and self-optimization, which logically follows the suppression of their pointless lazy days.

Considering these problems, it is reasonable to suggest that effective practices would include accepting some level of uselessness. This particular attitude does not promote remaining idle or cutting down on all the activities. Instead, it is more sympathetic to those modes of being, activities, situations, and ways of living that may seem uneconomical and impractical at first sight.

One such notion finds its conceptualization in the idea of "wu wei," a Chinese term that is translated as "non-doing," a common theme in Taoism and its practice. The essence of this principle is that relaxing and letting go are also useful and that some phenomena do not require struggle within their scope of possibilities. In more practical terms of modern life, this may

involve giving room to even boredom, spending time doing ‘nothing’, or ‘doing’ leisure where the aim itself is to not aim.

Embracing uselessness can have several potential benefits:

1. Increased mental wellness: Some unproductive time can be the right antidote to stress and anxiety and give the brain the rest it needs.
2. Boosted creativity: Downtime or participation in unimportant activities can be beneficial for creativity and idea generation. For example, neuroscientist Marcus Raichle has shown that the default mode network in our brains — which is functioning when we are not focused on a task — is fundamental for creativity and idea generation.
3. Reduced stress: Workers can endure healthy disruptions like non-productive time because it will assist them in creating appropriate divisions between their work and home space activities.
4. Higher satisfaction: People will perform activities of such a nature without anticipating any returns from it and just for the fun of it and they will get happier and content out of it.

On economic grounds, any instruction discouraging normal productive practices such as engaging in non-productive activities for some time might sound far-fetched. It could enhance the economy in the long run in several ways:

1. Avoiding burnout: This method may result in a more efficient labor force over time as it lessens the chances of employee burnout.
2. Encouraging creativity: Such idle periods may pave the way for unexpected lines of thought and new developments that spur the growth of an economy.
3. Boosting mental health: The population’s health improvement will have less healthcare costs and Great productivity overall will be generated.
4. Increase in flexibility: The workforce that is not completely occupied with generating returns on every task will possibly sustain better when confronted with economic changes and new technologies.

This two-sided strategy has been adopted in recent times by several liberal companies and societies such as. For example, some reproachful manner was practiced with technology companies using 20 percent rule internal policies where staff may work on any project for 20 percent of their work hours. Likewise, in Sweden to a certain extent, work days were also shortened because it was observed that increased hours of work did not translate to increased output.

Nevertheless, while productivity and usefulness will always remain the key issues of modern society, imposing a certain amount of uselessness as a necessary evil should be entertained. Through the adoption of certain moments of pause, meditation, or engaging in other activities that do not have an apparent desirable outcome, it is possible to further enhance the well-being and the economic sustainability of a society. Something that needs to change in this case is the hierarchy of needs of the population and the preferences of the individuals within the population in that not all activities need to be justified in terms of practical benefits. As it was

pointed out by the philosopher Bertrand Russell himself, there is nothing wrong with the virtue of idleness and he is against 'the idea that working is virtuous. This is all absurd. Everyone is doing a lot of harm in the modern world and believing that to be virtuous. Work is the key to happiness and prosperity but idleness is also necessary..

Conclusion:

I should say that consumer behavior focused on luxury products as part of consumer behavior is quite an interesting phenomenon that combines various aspects, such as psychological, social, and economic. Even though these kinds of goods should in many instances be classified as superfluous, which they are no doubt, they mostly remain the engine for economic growth and demand of households' consumption, but they also indicate the irrational aspects of a consumer where the consumers aim at status and social acceptance rather than functioning usefulness.

Even though the economy sometimes reaches a crisis, the luxury market never shows us anxiety rather only highlights what these products embody which is around the sickness of identity and a little bit about human exclusivity.

On the other hand, there are emerging concerns regarding this trend because of the growing amount of debt among consumers. The luxury market bubble is bound to explode as the market of counterfeits grows and consumers become all the more educated on the products in that market. The history of consumerism is about every human-over-generalizing their personality social credibility and appearance wherein even though the world will still forever love luxury – who wouldn't, the way people consume it is changing.

Heavy extension of the use of the average perceived value of luxury goods against the useable value, is not something expected in a utilitarian society. It brings forward other issues and at the same time answers the question of how seemingly useless spending can encourage productivity and positively satiate every individual. As society grapples with all these emerging issues, it is critical to appreciate the complex nature of consumerism and its associated effects, both on the individual and on the economy.

Works Cited

- “How inequality may be fuelling our obsession with luxury goods | News | Warwick Business School.” Warwick Business School, www.wbs.ac.uk/news/how-inequality-may-be-fuelling-our-obsession-with-luxury-goods.
- "When luxury democratizes: Exploring the effects of luxury" N.p., n.d. Web. 21 Aug. 2024 <<https://www.sciencedirect.com/science/article/abs/pii/S0148296322009134>>.
- Bochańczyk-Kupka, Dominika. (2019). NECESSITY AND LUXURY GOODS IN ECONOMICS. 10.32008/NORDSCI2019/B2/V2/16.
- Marples, Alice. “The South Sea Bubble of 1720 - The National Archives blog.” The National Archives Blog, 18 Sept. 2020, blog.nationalarchives.gov.uk/the-south-sea-bubble-of-1720.
- Resilience. “What is a ‘luxury economy,’ really?” Resilience.org, 18 Sept. 2023, www.resilience.org/stories/2023-09-18/what-is-a-luxury-economy-really.
- Schulz, Matt. “2024 Credit Card Debt Statistics.” LendingTree, 9 Aug. 2024, www.lendingtree.com/credit-cards/study/credit-card-debt-statistics/#:~:text=Americans%20have%20an%20absolute%20mountain,credit%20card%20use%20each%20month.
- Šliuburys, Vytautas. At the margins of productivity: philosophy of uselessness. Diss. 2017. The School of Life. “HISTORY OF IDEAS - Consumerism.” YouTube, 7 Oct. 2016, www.youtube.com/watch?v=Y-Unq3R--M0.
- Vincent Chan. “Luxury Brands Are For Broke People.” YouTube, 19 May 2024, www.youtube.com/watch?v=f0bIApjLBzA.
- What makes Modern Art so Expensive? | Live Your Passion. [live your passion.in/self-interest/art/Latest-trends-in-Art/What-makes-Modern-Art-so-Expensive--346](http://liveyourpassion.in/self-interest/art/Latest-trends-in-Art/What-makes-Modern-Art-so-Expensive--346).
- Zhan, May. “Does a Luxury Goods Presence Affect a Nation’s Happiness through the Lens of Income Inequality?” Bachelor of Science, 2012, www.stern.nyu.edu/sites/default/files/assets/documents/con_042983.pdf.
- "The Psychology Behind Why People Buy Luxury Goods." N.p., n.d. Web. 3 Sep. 2024 <<https://www.investopedia.com/articles/personal-finance/091115/psychology-behind-why-people-buy-luxury-goods.asp>>.

The Impact of Biological and Societal Factors on Individuals with ADHD

By Ava Bowles

Abstract

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder. Young adolescents between the ages of three and seven have a prevalence of 10.2% of being diagnosed with ADHD. Additionally, differences have been reported in the presentation of ADHD between the different genders, with the hyperactive/impulsive subtype more common in women and the inattentive subtype more common in men. The only consistent predictor of adult function in children with ADHD is childhood IQ. Additionally, this review article will cover the current standing of cognitive therapy by looking at the short and long-term effects and the side effects of stimulant medication. While ADHD is among the most commonly diagnosed and studied childhood psychiatric disorders, the developmental progress of individuals with ADHD is heterogeneous. This review article will cover a longitudinal study of ADHD, considering biological and societal factors that impact this disorder.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by a persistent pattern of inattention, hyperactivity/impulsivity, or a combination of the two that can interfere with functioning and development (Ringeisen et al.). Under the DSM-5 of ADHD, there are three diagnostic types: ADHD combined, ADHD impulsive/hyperactive, and ADHD inattentive (Kazda et al.). ADHD has a highly inheritable genetic origin, beginning in infancy and often continuing into adulthood (Pérez-Álvarez). The dopamine receptor D2 (DRD₂) gene regulates dopamine synthesis and storage, and its release is mutated in individuals with ADHD. This implicates neurons' ability to respond to dopamine, the neurotransmitter involved in feelings of pleasure and attention regulation (Blum et al.). This dysregulation of dopamine could explain the high frequency of individuals with ADHD being comorbid with other mood disorders, such as depression and anxiety.

Additionally, individuals with ADHD have low levels of norepinephrine, a neurotransmitter in the brain that plays an essential role in the regulation of attention and alertness, and low dopamine activity, which affects their cognitive processing and alters executive functions (Silver). Furthermore, research supports an increase in dopamine functioning in the nigrostriatal pathway (Fig 1) and a decrease in dopamine functioning in the mesocortical pathway (Castellanos). The increase in dopamine functioning in the nigrostriatal pathway (Fig 1) contributes to symptoms of impulsivity and hyperactivity, while the decrease in dopamine functioning in the mesocortical pathway (Fig 1) impacts individuals' selective attention ability (Engert and Pruessner). Besides the main behavioral symptoms of inattention, hyperactivity, and impulsivity, ADHD can also impair basic cognitive processes, resulting in slow processing speed, distractibility, and increased reaction time variability (Mohammad et al.).

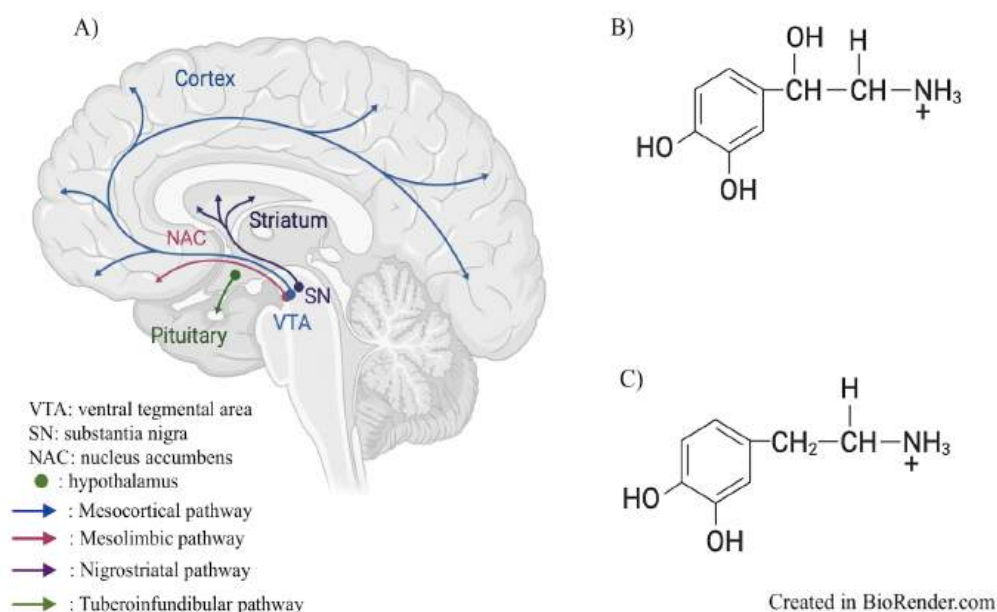


Fig 1: A) Dopaminergic pathways (left), B) norepinephrine chemical structure (upper right), and C) dopamine chemical structure (bottom right).

ADHD is one of the most common neurodevelopmental disorders, with about 11.3% of children in the US diagnosed between the ages of 5 and 17 (Reuben and Elgaddal). While ADHD is more prevalent in children, ADHD still affects more than 8 million adults in the US—about 5% of the population (Kessler et al.). Despite the majority of individuals with ADHD being children, it is important to analyze ADHD to fully understand predictors and outcomes longitudinally. To diagnose a child up to 16 years of age with ADHD, they must present up to six or more symptoms of inattention and hyperactivity-impulsivity. In contrast, individuals aged 17 and older must have five or more symptoms of inattention and/or hyperactivity-impulsivity (DSM-5). ADHD is currently one of the most commonly studied neurological disorders, and discoveries continue to surface. For example, researchers discovered there are different subtypes of ADHD, there are gender differences in the presentation of ADHD, and there are a number of innovative therapeutic and medical interventions for ADHD currently being tested (Reuben and Elgaddal; Miller; Law). With this wave of attention around ADHD, it is essential to take time to consider the impact of biological factors such as gender and age as well as societal factors, including IQ and SES, on individuals with ADHD.

The Different Types of ADHD

According to the DSM-5, there are three main subtypes of ADHD: ADHD impulsive/hyperactive, ADHD inattentive, and ADHD combined (Reuben and Elgaddal). In

1987, the DSM-3 was revised, and ADD was renamed ADHD after researchers observed no significant differences between hyperactive and non-hyperactive attention deficit disorder ADD (Epstein and Loren). Historically, researchers suggested the presence of three behavioral subtypes of ADD: without hyperactivity (ADD/WO), with hyperactivity (ADHD), and with hyperactivity and aggression (ADDHA) (Dykam and Ackerman). However, under the most recent classifications of ADHD, there are the three main subtypes mentioned above. It is vital to recognize the different subtypes because each subtype is distinct in symptomatology, clinical course and outcomes, neuro-psychologically, neuro-biologically, genetically, and in response to treatment (Showraki).

Subtype	Symptoms
ADHD hyperactive/impulsive	Squirms/fidgets when seated, appears to be driven by a motor, talks excessively, has difficulty waiting their turn, interrupts others, and impulsively blurts out
ADHD inattentive	Poor listening skills, loss of important items, easily distracted, short attention span, avoiding tasks requiring focus, and failure to pay attention to details
ADHD combined	Meets criteria for BOTH inattentive and hyperactive/impulsive subtypes: inability to pay attention, excessive talking, fidgeting, and making careless mistakes

Fig 2: Symptoms of each ADHD subtype

The first subtype is ADHD impulsive/hyperactive, which leans more towards the second part of ADHD—hyperactivity. Individuals under the ADHD impulsive/hyperactive subtype do not have trouble paying attention but are impulsive and hyperactive. The impulsive/hyperactive subtype is the least common subtype of ADHD, with only 7% of adults having the hyperactive/impulsive subtype (Wilens). As outlined by the DSM-5 Diagnostic Criteria for ADHD, the hyperactive symptoms of this subtype include “squirms when seated or fidgets with feet/hands, marked restlessness that is difficult to control, appears to be driven by a motor’ or is often ‘on the go,’ lacks the ability to play and engage in leisure activities in a quiet manner, incapable of staying seated in class, overly talkative”; the impulsive symptoms include: “difficulty waiting for turn, interrupts or intrudes into conversations and activities of others, and

impulsively blurts out answers before questions completed” (DSM-5). To be diagnosed with predominantly hyperactive/impulsive ADHD, the patient must meet the hyperactive/impulse criterion, but not the inattentive criterion, for the past 6 months (DSM-5).

Furthermore, the symptoms must be present before age 12 years, must not be better accounted for by a different psychiatric disorder (e.g., mood disorder, anxiety disorder), and must not occur exclusively during a psychotic disorder (e.g., schizophrenia), must not be exclusively a manifestation of oppositional behavior (DSM-5). As there are differences in behavioral outcomes of individuals among the three subtypes, this suggests there are also differences in brain structure among the three subtypes. Mu and colleagues recently found “dissociable structural brain changes across the different ADHD subtypes”. These structural brain differences among individuals between the three subtypes could explain the different behavioral outputs among the three subtypes. In one of Mu and colleague’s experiments, they found that the ADHD impulsive/hyperactive subtype had fewer symptoms of ADHD compared to ADHD combined and ADHD inattentive (Mu et al.; Milich et al.).

In addition to brain changes among the three subtypes, individuals with ADHD also have differences between their brain structure and a neurotypical individual’s brain structure (Wilens; Cronkleton). While behavioral differences between individuals with ADHD and neurotypical individuals are well known, the public less commonly understands the differences between someone with ADHD’s brain structure and a neurotypical person’s brain structure. Understanding that individuals with ADHD have tangible differences in the structure of their brain could help individuals with ADHD feel more validated in understanding their own behavior as well as reduce society’s stigma and stereotype of individuals faking or exaggerating ADHD symptoms (Corrigan and Shapiro).

The second subtype is ADHD inattentive, which includes individuals with higher inattention instead of hyperactivity/impulsivity. Furthermore, 31% of adults are a part of the inattentive subtype; this subtype is more common than the hyperactive/impulsive subtype but less common than the combined subtype (Wilens et al.). The symptoms of this subtype under the DSM-5 criteria include “displays poor listening skills, loses and/or misplaces items needed to complete activities or tasks, sidetracked by external or unimportant stimuli, forgets daily activities, diminished attention span, lacks the ability to complete schoolwork and other assignments or to follow instructions, avoids or is disinclined to begin homework or activities requiring concentration, and fails to focus on details and/or makes thoughtless mistakes in schoolwork or assignments.” In their study, Mu and colleagues found that ADHD-Inattentive patients had significantly reduced cortical thickness in the left posterior cingulate cortex and right lateral occipital lobe compared with the typically developing group (Mu et al.; Almeida et al.; Almeida et al.; Fortin et al.). Research has recognized that there is a pattern with reduced cortical thickness correlating with increased late-life depression and poorer response to antidepressant treatment response (Mackin). In addition, individuals with the ADHD inattentive subtype show milder clinical symptoms (ADHD Index and Hyper/Impulsive) compared to the ADHD combined group (Mu et al.; Milich et al.). These studies also concluded that individuals

with ADHD inattentive subtypes are more commonly comorbid with reading and language deficits compared to the other two ADHD subtypes (Mu et al.; Willcutt and Pennington). In Mu and colleagues' study, the ADHD-Inattentive patients had lower Verbal IQ than the ADHD-Combined patients, and Willcutt and Pennington's study also found that the association between reading disability and ADHD was stronger for symptoms of inattention than for symptoms of hyperactivity/impulsivity (Mu et al.; Willcutt and Pennington). In a review study of the inattentive and combined subtypes of ADHD examining patients' medical charts, researchers found that children with ADHD inattentive were more likely to be female, older (average age 12.4 compared to 11), have anxiety and depression disorders, have a history of speech/language therapy, and have a learning disability (Weiss et al.).

The final subtype of ADHD is the combined subtype. Individuals must meet both inattentive and hyperactive/impulsive criteria for the past six months to be diagnosed with combined ADHD (Ringeisen et al.). The combined subtype is the most common subtype of ADHD, with 62% of adults with ADHD reported to display this subtype (Wilens et al.). One experiment in Europe tested the persistence and outcomes of individuals with the ADHD combined subtype in late adolescence and young adulthood by measuring symptoms six years apart. Lieshout and colleagues found that although symptom severity decreased, persistence rates remained high, with 86.5% of participants maintaining a persistent DSM-5 ADHD diagnosis independent of age (Lieshout et al.). Although ADHD is commonly thought of as a childhood disorder, there is evidence that ADHD affects people of all ages. Mu and colleagues also found that patients with ADHD combined had significantly reduced cortical thickness, similar to the inattentive subtype (Mu et al.; Almeida et al.; Almeida et al.; Fortin et al.). However, one interesting difference they found was that children with ADHD combined tended to be extroverted, while children with ADHD-inattentive were more introverted (Mu et al.; Warner-Rogers et al.). When conducting a medical chart review study of the inattentive and combined subtypes of ADHD, Weiss and colleagues found that individuals with ADHD combined were more likely to have oppositional disorder, conduct disorder, been suspended/expelled, and been placed on and improved to stimulant medication (Weiss et al.).

How ADHD presents itself differently and changes in younger to older populations/ Correlation between ADHD predictors and neuropsychological outcomes

ADHD symptoms, predictors, and outcomes are all interrelated and affect an individual's life with ADHD. ADHD can start early in childhood and continue into adulthood for about a third of individuals (Low). ADHD can develop in adults or become apparent for the first time in adulthood (Fletcher). Although the DSM-5 states that symptoms of ADHD are the same for adults and children, there have been clinical observations suggesting that adults have more diverse deficits than children in higher-level executive functioning and emotional control (Adler et al.). Understandably, there are significantly different manifestations of ADHD in children than in adults due to maturation and development. However, many of the symptoms remain recognizably similar: children with ADHD inattentive squirming in cars and adults with ADHD

inattentive fidgeting during meetings. According to clinical psychologist Sybrand Hagan, inattention symptoms persist more into adulthood and show a slower decline than hyperactivity and impulsivity symptoms (Hagan).

Additionally, adults often internalize hyperactivity symptoms, which can be observed outwardly less often (Hagan). Adults with ADHD often struggle with occupational rank, academic achievement, and job performance as a result of the key symptoms of hyperactivity, inattentiveness, and impulsivity (Hagan). One study conducted in Italy found that patients with the hyperactive/impulsive subtype were proven to show worse quality of life and more frequent anxiety disorders (Salvi et al.). Of the many studies conducted to find a correlation between ADHD predictors and patient outcomes, it is clear that intelligence quotient (IQ) and socioeconomic status (SES) are the only two consistent predictors of ADHD outcome (Ramos-Olazagasti et al.; Cheung et al.). One study predicting the adult functional outcomes of boys with ADHD 33 years later found that childhood IQ was positively associated with educational attainment, occupational rank, and social and occupational adjustment (Ramos-Olazagasti et al.). It is important to note that IQ can be considered a problematic or biased measure as IQ is largely affected by SES; higher SES increases access to education, contributing to a higher IQ. In addition, SES and reading ability were positively associated with educational attainment (Ramos-Olazagasti et al.; Mannuzza et al.; Grevin et al.). Ramos-Olazagasti and colleagues' study found that the four significant adolescent predictors of ADHD were: antisocial behaviors, which predicted poorer educational attainment, educational goals, which related to better overall function, early job functioning, which had a positive relation with social functioning, and early social functioning which was positively related to occupational functioning. Overall, Ramos-Olazagasti and colleagues' study found no consistent predictors of adult function among children with ADHD besides childhood IQ (Ramos-Olazagasti et al.; Klein et al.; Cheung et al.).

Supporting Ramos-Olazagasti and colleagues' study, Cheung and colleagues found that SES and IQ were significant predictors of ADHD symptoms and outcome, childhood severity of ADHD symptoms predict later ADHD outcomes, and no other childhood cognitive measures besides IQ predict later ADHD severity (Cheung et al.). This same study found that other psychiatric comorbidities, oppositional defiant disorder, anxiety disorder, and family factors, including maternal psychopathology and psychosocial adversity, significantly predicted the persistence of ADHD in adolescence and adulthood (Cheung et al.; Reale et al.). Lastly, a research study on the neuropsychological outcome in adolescents/young adults with childhood ADHD, testing groups of persisters, remitters, and controls, found that those with childhood ADHD had more broad neuropsychological deficits compared to those in the control group (Halperin et al.; Caye et al.). Furthermore, the ADHD persisters and remitters showed deficiency in "perceptual sensitivity" and "response variability," which supports that ADHD is associated with early-appearing symptoms and that recovery over time is associated with improvements in executive control functions (Halperin et al.; Campbell et al.).

In conclusion, the presentation and progression of ADHD throughout an individual's lifespan presents high variability, with differences in the manifestation of symptoms emerging from childhood to adulthood. The expression of the core symptoms, including inattention, hyperactivity, and impulsivity, persist but are influenced by developmental factors and environmental factors. Evident in this section, SES and IQ are both significant predictors of long-term outcomes for individuals with ADHD, however, this correlation is complex, and the two even affect each other, with SES and IQ often being interlinked. Overall, understanding how ADHD presents itself differently in different demographics and how predictors correlate with outcomes is crucial for designing early interventions and improving the quality of life for those with ADHD.

Gender differences

Traditionally, gender differences in ADHD have been mistaken or inaccurate due to small sample sizes in research studies and increased research on men. To start, one major difference between men and women with ADHD is that men tend to externalize their symptoms more often, which makes it easier to recognize and diagnose ADHD in men (Miller). Women tend to internalize symptoms which often makes it hard for them to get the help they need (Miller). This difference between internalizing and externalizing symptoms in men and women has previously led to inaccurate data.

Furthermore, researchers have shown that men and women carry symptom differences as well. One study found that girls with ADHD had lower ratings on hyperactivity, impulsivity, and externalizing problems, as well as more significant intellectual impairments, compared to boys with ADHD (Skogli et al.). Moreover, adolescent girls are more often diagnosed as predominantly inattentive, have lower self-efficacy, poorer coping strategies, higher rates of depression and anxiety, and lower physical aggression than adolescent boys (Gershin et al.; Rucklidge). This introduces a common finding that women have higher comorbidity of mood disorders such as depression and anxiety (Gershin et al.; Ottosen et al.; Quinn,; Biederman et al.; Tung et al.). One research study evaluating the socio-demographic and clinical characteristics of a sample of adults with ADHD and the characteristics associated with the different disorder subtypes found that the hyperactive/impulsive subtype was significantly more frequent in women, and the inattentive subtype was more frequent in men (Salvi et al.). Salvi and colleagues' study produced findings that reiterated Rucklidge and Gershin's previous conclusions, emphasizing women's predominant traits.

In an analysis of sex differences in human brain structure, focusing on volume, Ruigrok and colleagues found that regional sex differences correlate with areas affected by psychiatric conditions. For example, the amygdala, insula, planum temporale, and hippocampus all present sex differences (Ruigrok et al.). This is significant information since the amygdala is responsible for emotional response (Johns), the insula plays a role in the generation of emotions (Helion and Ochsner), the planum temporale is in charge of auditory processing (McCarly and Shenton), and the hippocampus stores information including memories, emotions, and learning (Zhu and Shen).

Therefore, the regional sex differences help explain gender differences in the presentation of ADHD as an emotional response, sensory processing, and information storage are all impacted and vary by sex (Ruigrok et al.; Salvi V et al.; Rucklidge; Gershin et al.). For example, the average size of a male's left amygdala, the region in charge of emotional processing, is more significant than that of a female's, which supports why males have, on average, better emotional responses with less comorbidity and better coping strategies than females (Baeken).

A systematic literature review to examine the sex differences in prescription rates and efficacy of pharmacotherapy treatment in girls and women with ADHD found that girls received significantly fewer prescriptions than boys; however, this was not the case for adults (Kok et al.). Additionally, each of the 14 studies on effectiveness reviewed by Kok and her colleagues found at least one sex difference in the effects of ADHD pharmacotherapy (Kok et al.). The presence of sex differences in the effects of ADHD pharmacotherapy indicates that the gender differences of ADHD, starting with structural brain differences, are significant enough that there are also differences in medication depending on an individual's gender.

The differences in behavior between genders also reflect differences in society's expectations of men and boys versus women and girls. In addition to brain differences, society's expectations have a huge impact on the mindset and actions of each gender. For example, women tending to internalize their symptoms of ADHD is a major reflection of society's expectation for women to be submissive and not make a scene (Thurber et al.; Diamantopoulou et al.). Society's standards for men and boys are the opposite, allowing males to be impulsive and display hyperactive behavior (Thurber et al.; Diamantopoulou et al.). This may contribute to why males present higher numbers of the hyperactive and impulsive subtypes and externalize their problems more frequently.

Cognitive Function, Impact, and Comorbidity

Firstly, it is important to understand the basic cognitive functions of individuals with ADHD to understand deficits, strengths, and the impact of ADHD on an individual's quality of life. Basic cognitive functions include sensory processing and perception, attention, motor skills, and language (Attwood et al.). Complex cognitive functions include long-term memory, decision-making, problem-solving, and critical thinking (Zhang). One study compared the basic cognitive functions like processing speed and distractibility and more complex functions like working memory and cognitive flexibility in adults with ADHD to adults without ADHD (Mohamed et al.). This study found that individuals with ADHD were slower in processing speed, had more problems with selective attention, and had worse memory than those without ADHD—suggesting evident issues with cognitive functions (Mohamed et al.). Furthermore, executive functions explained 25% of the differences between adults with and without ADHD, and basic functions like processing speed and distractibility explained about 64% of problems with complex functions (Mohamed et al.). This shows that issues with basic cognitive skills can have a more significant impact on more complex functions like thinking and memory.

Additionally, one further challenge individuals with ADHD face is having comorbid mental health challenges, commonly including mood disorders like anxiety and depression (Hou et al.; Reale et al.; Jensen et al.). Hou and colleagues state, "...one in 50 (2.1%) children with a diagnosis of ADHD also have a mood disorder, such as depression, while more than one in four (27.4%) have an anxiety disorder" (Hou et al.). One additional study on comorbidity prevalence and treatment outcome in children and adolescents with ADHD found that 66% of patients who met diagnostic criteria for ADHD had at least one comorbid psychiatric disorder such as learning disorder, sleep disorder, oppositional defiant disorder, or anxiety disorder (Reale et al.). Furthermore, patients with the combined subtype of ADHD were more likely to have comorbidity (Reale et al.; Jensen et al.). In this study, 67% of the subjects evaluated met the diagnostic criteria for ADHD, 34% had only ADHD, and 66% had at least one comorbid psychiatric disorder. Of the individuals with a comorbid psychiatric disorder, 56% had a learning disorder, 23% had a sleep disorder, 20% had oppositional defiant disorder, and 12% had an anxiety disorder (Reale et al.). This shows that having ADHD comes with further mental and emotional difficulties and differing neurological processes than the standard ADHD symptoms suggest. It is important to keep this high rate of comorbidity of individuals with ADHD in mind, to best understand how to support and treat individuals with ADHD.

One research study assessed whether ADHD children with and without comorbid conditions have equally high levels of core symptoms and whether symptom profiles differ as a function of comorbidity and gender (Newcorn et al.; Hou et al.). This study found that children with ADHD had high levels of inattention, impulsivity, and dyscontrol on a continuous performance test (CPT), regardless of comorbidity (Newcorn et al.). However, certain trends were present, including children with ADHD and OCD being more impulsive than other ADHD subgroups and children with ADHD and anxiety disorders being relatively more inattentive than impulsive (Newcorn et al.; Reale et al.). Additionally, girls with ADHD and anxiety made fewer impulsivity errors on the CPT than girls with ADHD-only—children with ADHD and anxiety were more prominent in presenting as inattentive rather than hyperactivity/impulsivity (Newcorn et al.). Overall, symptom profiles differed depending on the specific comorbidities present, and core ADHD symptoms varied as a function of both comorbidity and gender (Newcorn et al.). This finding is crucial as it supports that individuals being diagnosed with ADHD should also be tested for comorbid conditions to get a comprehensive diagnosis and plan of action on how to best treat their specific comorbidities since this affects symptom profiles.

Besides comorbid conditions, another related aspect to consider is how ADHD often increases social and emotional challenges. For example, in Classi and colleagues' study on social and emotional difficulties in children with ADHD and the impact on school attendance and healthcare utilization, they found that the presence of social and emotional problems resulted in greater odds of missed school days and healthcare utilization (Classi et al.). Emotional difficulties often include poor emotional self-regulation, aggression, and reduced empathy (Classi et al.; Bodalski et al.). These traits commonly manifest in social difficulties, leading to conflicts with family and problems with peers (Classi et al.; Rich et al.). Individuals with ADHD

may have poorer emotional self-regulation and aggression explaining why increased conflict leading to suspension or probation as well as decreased school attendance is more common.

Emotional dysregulation, defined by three subscales of temper control, affective lability (the tendency to experience rapid, excessive, and unpredictable changes in affective states), and emotional over-reactivity (stress intolerance), is highly prevalent in individuals with ADHD (Corbisiero et al.; Høegh et al.). Corbisiero and colleagues inquired whether emotional dysregulation is part of the psychopathology of ADHD in adults and found that emotional dysregulation symptoms are as prevalent as the classic ADHD symptoms of inattention, hyperactivity, and impulsivity in adult ADHD samples (Corbisiero et al.). Furthermore, emotional dysregulation significantly contributes to impairments in major life activities beyond what is caused by classic ADHD symptoms alone (Corbisiero et al.; Bodalski et al.). Just like classic ADHD symptoms, emotional dysregulation symptoms respond to stimulant medications like methylphenidate (MPH) and atomoxetine (ATX) (Corbisiero et al.). Overall, symptoms of emotional dysregulation are definable and appear to be distinct factors of the psychopathology of adult ADHD.

A study by Butzbach and colleagues explored metacognition (thinking about one's thinking), specifically self-awareness of cognitive performance, in adults with ADHD compared to controls (Butzbach et al.). In this study, the test group had ADHD combined, and the control group did not have ADHD (Butzbach et al.). To assess metacognition, participants evaluated their own test performance after completing neuropsychological tests, and then the discrepancy between self-evaluation and actual performance was calculated (Butzbach et al.). To start, adults with ADHD showed impairments in attention and memory on a neuropsychological test, but no significant deficits in executive functions (Butzbach et al.; Thaler et al.). Furthermore, adults with ADHD showed a deficit in self-awareness of attentional functions, overestimating their performance on attention tasks compared to controls (Butzbach et al.; Kofler et al.). However, subjective reports (self and informant) of cognitive functioning did not reveal differences between patients and controls, unlike the objective measure of discrepancy scores (Butzbach et al.). This study supports that metacognitive deficits exist in adults with ADHD, specifically reduced self-awareness of attention (Butzbach et al.; Alvarado et al.; Poissant; Butzbach et al.).

Therapeutic and Medical Intervention

There are many different types of medication for individuals with ADHD, but the main types are stimulants and non-stimulants, and sometimes antidepressants (Law). ADHD medications increase neurotransmitters, including dopamine and norepinephrine (Fig 1), in an individual's brain (Law). Also, the medication for ADHD is important as it helps individuals with ADHD focus their thoughts, ignore distractions, pay attention, and control their behavior (Holden).

The first type of medication for ADHD, stimulants, is the most common form of medication for ADHD. Stimulants are controlled substances, which means their production and distribution are highly regulated since the medication may be abused or cause addiction

(National Cancer Institute). There are two types of stimulants: immediate-release and extended-release (Greenhill et al.). Immediate release stimulants are short-acting and last up to four hours (Greenhill et al.). Extended release is intermediate or long-acting and lasts six to eight hours or up to 16 hours (Greenhill et al.). Stimulant medication is effective in treating ADHD as it increases dopamine and norepinephrine in an individual's brain (Law). An estimated 80% of children with ADHD have fewer symptoms after finding the correct stimulant medication type and dosage (Advokat).

Next, non-stimulants are the second type of medication for ADHD (Law). In contrast to stimulant medication, non-stimulants are non-controlled substances (Budur et al.). Hence, non-stimulant medication is less likely to be overused, and an individual is less likely to become dependent (Budur et al.). Non-stimulant medication increases levels of norepinephrine and dopamine in the cortex (Budur et al.). However, non-stimulant medication takes longer than stimulant medication to start working and can work up to 24 hours (Cleveland Clinic). Non-stimulants only affect the cortical regions, unlike stimulants which only affect the subcortical and cortical regions of the brain (Law). A common non-stimulant medication is atomoxetine which increases dopamine levels in the synaptic cleft, increasing activity in the prefrontal cortex, the ventral tegmental area, and the nucleus accumbens (Guzman; Rawson). Increasing activity in these regions leads to higher self-regulation and executive functions, including planning, attention, and behavior (Law).

Lastly, antidepressants are another medication that some healthcare providers prescribe to patients with ADHD; however, the FDA has not specifically approved the use of antidepressants for ADHD (Otasowie et al.). Antidepressants can be taken alone or with a stimulant. Taking antidepressants may result in slight sedation, so, if taken with a controlled substance like a stimulant, including amphetamines, it could counteract the sedating side effect (Anderson). Like the other two types of medication for ADHD, antidepressants work on the dopamine and norepinephrine levels in your brain. One study on ADHD treatment in primary care, looking at demographic factors, medication trends, and treatment predictors, found that predictors of antidepressant prescriptions included increasing age, psychiatric consultation, and diagnoses of both anxiety and depression (Hauck et al.). Moreover, in Hauck and colleagues' study of ADHD patients taking antidepressants, "7.5% had received a prescription for bupropion (further breakdown is not possible because the number of patients becomes too small)" (Hauck et al.).

Taking ADHD medications may also have side effects during early treatment. Decreased appetite is one of the most common side effects of ADHD medication, affecting around 80% of people who take stimulant medication (Law). Weight loss is another common side effect of taking ADHD medication; this side effect can be prevented by eating before taking medication (Law). Difficulty sleeping is the last of the three most common side effects where individuals take longer to fall asleep or have poorer quality of sleep (Cascade et al.). Other side effects from taking ADHD medication include the rebound effect of fatigue (increased activity or mood changes as the medication wears off), anxiety or increased depression, tics like eye blinking or throat clearing, minor growth delay for children/teens, upset stomach with nausea and vomiting,

changes in blood pressure, increased heart rate (Cascade et al.; Toomey et al.). If side effects last more than a couple of weeks, it may be good for the individual to change the dosage of the medication prescribed, use a different stimulant, or change their schedule (Cascade et al.).

Besides medication, behavioral therapy is another effective treatment to help individuals with ADHD. A review study on the interventions in ADHD comparing stimulant medication and behavioral therapies found that behavioral interventions like cognitive behavioral therapy are promising for improving functional, academic, and social outcomes in the long term (Rajeh et al.). However, in this study, behavioral therapies were less effective than stimulants for core ADHD symptoms (Rajeh et al.). Additionally, combination therapy with stimulants and behavioral interventions indicates greater benefits for functioning than each therapy alone (Rajeh et al.). Rajeh and colleagues also found that stimulant medications like methylphenidate and amphetamines are effective in controlling ADHD symptoms in the short term. However, their benefits in the long term are still unclear (Rajeh et al.).

Cognitive function occupational therapy treats individuals with ADHD since it addresses executive functioning, focusing on behaviors, the environment, thoughts, and emotions (Gharebaghy et al.). According to the School of Occupational Therapy, cognitive functional therapy aims to “promote quality of life in persons with ADHD and their families, improve participation in daily occupations, acquire executive strategies to compensate for typical ADHD difficulties in daily life, promote awareness to ADHD symptoms and implications through education and experiential learning, enable successful experiences and feelings of efficacy in challenging daily activities, and educate and empower parents as mediators in the therapeutic process”. One research study on the efficacy of Cognitive–Functional (Cog–Fun) occupational therapy intervention among children with ADHD found that Cog-Fun intervention in occupational therapy shows positive effects for improving executive functions, quality of life, and decreasing symptoms among children with ADHD (Hahn-Markowitz et al.). Additionally, a separate study on the effectiveness of the Cog-Fun intervention for young children with ADHD also found that Cog–Fun intervention successfully improved occupational performance and executive functions in daily life (Maeir et al.).

Conclusion

In summary, Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder influenced by both biological and environmental factors (Freitag et al.). As stated in this review, ADHD is caused by reduced dopamine and norepinephrine activity, mainly in the prefrontal cortex, affecting people’s cognitive processing and executive functions (Silver; Del Campo et al.; Madras et al.; Mehta et al.). The three subtypes of ADHD include: ADHD-combined, ADHD-inattentive, and ADHD-hyperactive/impulsive—in order of most common to least common (DSM-5; Wilens et al.). The three subtypes differ in symptom profiles and outcomes: women more commonly have the inattentive subtype, individuals with the impulsive/hyperactive subtype, on average, had fewer symptoms, and people with the combined subtype were found more likely to have oppositional disorder (Weiss et al.; Mu et al.).

Furthermore, while genetics play a role in ADHD heredity, the structural differences between male and female brains contribute to gender differences in the presentation of ADHD (Ruigrok et al.; Salvi V et al.; Rucklidge; Gershin et al.; Ruigrok et al.). It is important to recognize that women often go undiagnosed because they tend to have inattentive symptoms, internalize their symptoms, and have comorbid disorders like depression and anxiety (Miller; Gershin et al.). In addition to gender having an impact on presentation severity and outcome, childhood predictors like lower IQ and lower socioeconomic status are correlated with poorer academic achievement, more occupational difficulties, and social impairments later in life (Olazagasti et al.; Cheung et al.).

Another main takeaway is that ADHD symptoms impact executive functions, commonly resulting in slower processing speed, increased problems with selective attention, and worse working memory (Mohammad et al.). On top of impacted executive functions, individuals with ADHD are also more susceptible to having comorbid mental health challenges, commonly including mood disorders like anxiety and depression (Hou et al.; Reale et al.). ADHD can also have social and emotional effects, with emotional difficulties commonly including poor emotional self-regulation, aggression, and reduced empathy, and social difficulties commonly including conflicts with family and problems with peers (Classi et al.). However, there are a range of medications available for individuals with ADHD. The most common types are stimulants, non-stimulants, and antidepressants which work by stabilizing the levels of dopamine and norepinephrine in the brain (Law). Medications are usually successful in reducing symptoms; however, cognitive function occupational therapy and behavioral therapy are two additional long-term aids that help improve impairments in executive function (Rajeh et al.; Hahn-Markowitz et al.; Maeir et al.). Exploring whether an increase in technology correlates with or has contributed to the increase in the prevalence of ADHD would be an interesting future area of study. Overall, further research would still be beneficial in better understanding gender differences in treatment response due to the longstanding increased focus on men and misunderstanding of female presentation and the development of more personalized combination therapy approaches across an individual's lifespan to help individuals with ADHD overcome difficulties (Miller; Rajeh et al.).

Works Cited

- Alder, Lenard A., et al. "The structure of adult ADHD." Wiley Online Library, 27 November 2016, <https://doi.org/10.1002/mpr.1555>.
- American Psychiatric Association. "Diagnostic and Statistical Manual of Mental Disorders, 5th edition, text revision." Psychiatry Online, 2022.
- "Attention-Deficit / Hyperactivity Disorder (ADHD) in Children." Johns Hopkins Medicine, <http://www.hopkinsmedicine.org/health/conditions-and-diseases/adhdadd>.
- "Attention-Deficit/Hyperactivity Disorder (ADHD) - National Institute of Mental Health (NIMH)." National Institute of Mental Health, <http://www.nimh.nih.gov/health/statistics/attention-deficit-hyperactivity-disorder-adhd>.
- B, Gaynes, et al. "Future Research Needs—ADHD Research Needs - Future Research Needs for Attention Deficit Hyperactivity Disorder." NCBI, May 2021, <https://www.ncbi.nlm.nih.gov/books/NBK83964/>.
- Blum, Kenneth. "'Attention-Deficit/Hyperactivity Disorder and Reward Deficiency Syndrome.'" National Center for Biotechnology Information, U.S. National Library of Medicine, 1 January 1970, www.ncbi.nlm.nih.gov/pmc/articles/PMC2626918/.
- Budur, Kumar, et al. "Non-Stimulant Treatment for Attention Deficit Hyperactivity Disorder." NCBI, July 2005, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000197/>.
- Butzbach, Marah, et al. "Metacognition in adult ADHD: subjective and objective perspectives on self-awareness of cognitive functioning." PubMed, 19 January 2021, <https://pubmed.ncbi.nlm.nih.gov/33464422/>.
- Cascade, Elisa, et al. "Real-World Data on: Attention Deficit Hyperactivity Disorder Medication Side Effects." PubMed, <https://pubmed.ncbi.nlm.nih.gov/20508803/>.
- Cheung, Celeste, et al. "Childhood predictors of adolescent and young adult outcome in ADHD." *National Center for Biotechnology Information*, PubMed, 29 January 2015, <https://pubmed.ncbi.nlm.nih.gov/25680235/>.
- Classi, Peter, et al. "Social and emotional difficulties in children with ADHD and the impact on school attendance and healthcare utilization." *NCBI*, PubMed Central, 4 October 2012, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3489829/>.
- Corbisiero, Salvatore, et al. "Is emotional dysregulation part of the psychopathology of ADHD in adults?" PubMed, 5 June 2013, <https://pubmed.ncbi.nlm.nih.gov/23208078/>.
- Dykman, Roscoe A., and Peggy Ackerman. "Behavioral Subtypes of Attention Deficit Disorder." *Exceptional Children*, Sage Journals, October 1993, <https://journals.sagepub.com/doi/10.1177/001440299306000207>.
- Engert, Veronika, and Jens C. Pruessner. "Dopaminergic and noradrenergic contributions to functionality in ADHD: the role of methylphenidate." *Current Neuropharmacology*, 2008.
- Epstein, Jeffery N., and Richard EA Loren. "Changes in the Definition of ADHD in DSM-5: Subtle but Important." *NCBI*, PubMed Central, 1 October 2013, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3955126/>.
- "The Four Causes of ADHD: Aristotle in the Classroom." *Frontiers*, 8 June 2017, <https://doi.org/10.3389/fpsyg.2017.00928>.

- Freitag, Christine M., et al. "Biological and psychosocial environmental risk factors influence symptom severity and psychiatric comorbidity in children with ADHD." *Journal of Neural Transmission*, 31 May 2011, <https://link.springer.com/article/10.1007/s00702-011-0659-9>.
- Greenhill, Laurence L., et al. "Stimulant Medications." *Journal of the American Academy of Child & Adolescent Psychiatry*, ScienceDirect, May 1999, <https://www.sciencedirect.com/science/article/pii/S0890856709631312>.
- Hagan, Sybrand. "The difference between child and adult attention-deficit/hyperactivity disorder." *Mental Health Matters*, 2019, https://www.sadag.org/images/pdf/MHM_Difference-between-Child-and-adult-ADHD.pdf.
- Halperin, Jeffrey M., et al. "Neuropsychological Outcome in Adolescents/Young Adults with Childhood ADHD: Profiles of Persisters, Remitters and Controls." *National Center for Biotechnology Information*, PubMed Central, 28 June 2008, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2646044/>.
- Hauck, Tanya S., et al. "ADHD Treatment in Primary Care: Demographic Factors, Medication Trends, and Treatment Predictors." PubMed, 19 January 2017, <https://pubmed.ncbi.nlm.nih.gov/28103079/>.
- Høegh, Margrethe C., et al. "Affective lability and social functioning in severe mental disorders." *National Library of Medicine*, PubMed, 27 January 2022, <https://pubmed.ncbi.nlm.nih.gov/35084540/>.
- Hou, Wenjie, et al. "Emotion dysregulation and right pars orbitalis constitute a neuropsychological pathway to attention deficit hyperactivity disorder." *nature mental health*, 13 May 2024, <https://www.nature.com/articles/s44220-024-00251-z>.
- Kessler, Ronald C., et al. "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication." *PubMed*, June 2005, <https://pubmed.ncbi.nlm.nih.gov/15939837/>.
- Kok, Francien M., et al. "The female side of pharmacotherapy for ADHD—A systematic literature review." *Plos*, 18 September 2020, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239257>.
- Lieshout, Marloes van, et al. "A 6-year follow-up of a large European cohort of children with attention-deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood." *European Child & Adolescent Psychiatry*, Springer Link, 2 February 2016, <https://link.springer.com/article/10.1007/s00787-016-0820-y>.
- Mackin, Scott, et al. "Patterns of Reduced Cortical Thickness in Late Life Depression and Relationship to Psychotherapeutic Response." *National Center for Biotechnology Information*, PubMed Central, 6 February 2013, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3732520/>.
- Maeir, Adina, et al. "Effectiveness of Cognitive-Functional (Cog-Fun) occupational therapy intervention for young children with attention deficit hyperactivity disorder: a controlled study." PubMed, May-June 2014, <https://pubmed.ncbi.nlm.nih.gov/24797189/>.
- Mohamed, Saleh, et al. "Basic and complex cognitive functions in Adult ADHD." *NCBI*, PubMed Central, 2 September 2021, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8412315/>.
- Newcorn, Jeffery. "Symptom profiles in children with ADHD: effects of comorbidity and gender." PubMed, February 2001, <https://pubmed.ncbi.nlm.nih.gov/11214601/>.

- Rajeh, Adnan, and Shabbir Amanullah. "Interventions in ADHD: A comparative review of stimulant medications and behavioral therapies." *Asian Journal of Psychiatry*, 1 September 2016, <https://bpb-us-w2.wpmucdn.com/hawksites.newpaltz.edu/dist/f/840/files/2016/11/Interventions-in-ADHD-27c1618.pdf>.
- Ramos-Olazagasti, María A., et al. "Predicting the Adult Functional Outcomes of Boys With ADHD 33 Years Later." *Journal of the American Academy of Child & Adolescent Psychiatry*, ScienceDirect, August 2018, <https://www.sciencedirect.com/science/article/abs/pii/S0890856718302806>.
- Reale, Laura, et al. "Comorbidity prevalence and treatment outcome in children and adolescents with ADHD." *European Child & Adolescent Psychiatry*, Springer, 19 May 2017, <https://link.springer.com/article/10.1007/s00787-017-1005-z>.
- Reuben, Cynthia, and Nazik Elgaddal. "Products - Data Briefs - Number 499 - March 2024." *CDC*, 20 March 2024, <http://www.cdc.gov/nchs/products/databriefs/db499.htm>.
- Rucklidge, Julia J. "Gender differences in ADHD: Implications for psychosocial treatments." *Expert Review of Neurotherapeutics*, ResearchGate, May 2008, https://www.researchgate.net/profile/Julia-Rucklidge/publication/5435322_Gender_differences_in_ADHD_Implications_for_psychosocial_treatments/links/54c6d22c0cf289f0cecc5bcb/Gender-differences-in-ADHD-Implications-for-psychosocial-treatments.pdf.
- Ruigrok, Amber, et al. "A meta-analysis of sex differences in human brain structure." *Neuroscience & Biobehavioral Reviews*, ScienceDirect, February 2014, <https://www.sciencedirect.com/science/article/pii/S0149763413003011>.
- Sainburg, Lucas E., et al. "Characterization of resting functional MRI activity alterations across epileptic foci and networks." *Cerebral Cortex*, Oxford Academic, 15 December 2022, <https://academic.oup.com/cercor/article-abstract/32/24/5555/6527282?redirectedFrom=fulltext>.
- Saleh, Mohamed. "Basic and complex cognitive functions in Adult ADHD." *National Center for Biotechnology Information*, National Library of Medicine, 22 September 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8412315/.
- Salvi, Virginio, et al. "ADHD in adults: clinical subtypes and associated characteristics." *Rivista di psichiatria*, 2019, <https://www.rivistadipsichiatria.it/archivio/3142/articoli/31249/>.
- Silver, Larry. "The ADHD Brain: Neuroscience Behind Attention Deficit Disorder." *ADDitude*, 8 April 2024, <http://www.additudemag.com/neuroscience-of-adhd-brain/>.
- Substance Abuse and Mental Health Services Administration. "DSM-5 Changes." *Implications for Child Serious Emotional Disturbance*, June 2016, <https://www.ncbi.nlm.nih.gov/books/NBK519708/>.
- Tripp, Gail, et al. "Parent and teacher rating scales in the evaluation of attention-deficit hyperactivity disorder: contribution to diagnosis and differential diagnosis in clinically referred children." *PubMed*, 27 June 2006, <https://pubmed.ncbi.nlm.nih.gov/16775518/>.
- University of Cambridge. "One in two children with ADHD experience emotional problems, study finds." 22 May 2024, <https://www.cam.ac.uk/research/news/one-in-two-children-with-adhd-experience-emotional-problems-study-finds#:~:text=>
- Weiss, Margaret, et al. "A Chart Review Study of the Inattentive and Combined Types of ADHD." *Journal of Attention Disorders*, ResearchGate, October 2003.
- Wilens, Timothy E., et al. "Presenting ADHD Symptoms, Subtypes, and Comorbid Disorders in Clinically Referred Adults with ADHD." *NCBI*, PubMed Central, November 2009, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2948439/>.

Comparing Precision of Convolutional Neural Networks (CNN) Model on Audio Deepfake Detection Across English and Russian Languages By Saahil Manekar

Abstract

With the rise in technology and Artificial Intelligence, DeepFakes are being produced rapidly. DeepFakes are presented as videos, photos, and audios, causing the spread of misinformation and inflicting global harm. Currently, many researchers and Artificial Intelligence experts are developing deepfake detectors but face challenges such as model overfitting, manual preprocessing, and most notably, adaptability to languages besides English. In this research paper, we test the extent to which a Convolutional Neural Network (CNN) model generalizes across English and Russian languages by comparing the precision of the two datasets: the English audio dataset and the Russian audio dataset. Through extensive preprocessing, we convert audio files into waveforms and mel spectrograms, preparing the audio data for model training and testing. After training and testing our CNN model on the English and Russian datasets, we observe a wide difference between the results, including precision, recall, loss, and training and testing time. We believe that our research encourages future researchers to create more robust machine learning models to classify audio deepfakes and real audios precisely and address the limitations we face during the research process.

Introduction

Deepfakes are recognized today as synthetic media, including pictures, videos, and audios, generated by Artificial Intelligence and machine learning techniques. The term “deepfake” emerges in 2017 when a Reddit user swaps a celebrity's face onto another, and ever since Deepfakes have become a threat to society. With countless other online platforms and applications enabling users to continue spreading Deepfakes, such as DeepFaceLab and FakeApp, over 15,000 deepfake videos exist today (George and George 58-59).

Deepfakes are created via generative adversarial networks (GANs), which consist of a generator and a discriminator. The generator creates fake media, while the discriminator attempts to identify whether the media is indeed fake. Through the use of large amounts of data, the generator eventually creates media that the discriminator cannot correctly detect as fake (George and George 59).

With the increasing popularity and advancements of Artificial Intelligence, Deepfakes become a more pressing issue, introducing various challenges: spreading fake news, misleading military personnel, creating financial fraud, and causing political confusion (Mahmud and Sharmin 13)

Additionally, Deepfakes spread globally, endangering a large audience through misinformation. In most cases, current studies and strategies to surmount the challenge of Deepfake audio detection rely on English audios which do not properly represent our world today with thousands of languages. The Arabic language, for example, is the fourth most

commonly spoken language; however its unique alphabet pronunciation makes it difficult for machine learning models to correctly identify Deepfake audios (Almutairi and Elgibreen 16)

With Deepfakes in multiple languages emerging as a global issue, this research paper aims to determine whether a Convolutional Neural Network (CNN) model can generalize across different dialects: English and Russian. By comparing the precision of the CNN model in a multilingual context, we can evaluate how effectively the model classifies audio deepfakes in the real world.

In this paper, we build the same CNN model for each language, meaning that each model for each language is preprocessed, trained, and tested using the same techniques, audio length, and parameters to avoid bias. This approach ensures that a certain model's precision does not differ from another due to the amount of input data or a change in the model's structure.

When machine learning models' content differs, the discrepancy between the two precisions can result from differences in model content, biasing the actual performance of the model. This falsified result leads to misleading comparisons between languages. For example, we might conclude that the CNN model generalizes well if its precision is similar in both Russian and English, while in reality, the precisions are far off meaning the CNN model doesn't generalize effectively.

Instead we aim to address practical reasons that stem from the difference between precisions, such as variations in audio quality or pronunciation of complex words in a language. These reasons provide a clearer understanding of how well the model generalizes in a multilingual context and help form more accurate evaluations. In addition, future research can focus on tackling these issues instead of merely working to eliminate the clear bias of differences in model content or input data.

Background

A Convolutional Neural Network (CNN) contains an intricate architecture of multiple layers and serves as a powerful classification tool. The structure of a CNN is inspired by the neurons in our brains. CNNs can identify and draw distinct patterns from the data without human supervision, enabling them to classify accurately. Since most CNN models are trained on images, the following description explains how a CNN model works in the context of images (Alzubaidi et al. 2)

The CNN architecture consists of four distinct layers: convolution, pooling, fully connected layer, and non-linearity layers. When viewing an image, the machine examines the pixels or numbers that make up the image. The convolution layer's kernel filter, which matches the dimension of the input image, obtains unique characteristics of the image as the filter mask slides across it. A product is calculated between the pixels and the kernel filter weight, resulting in a 2D activation map.

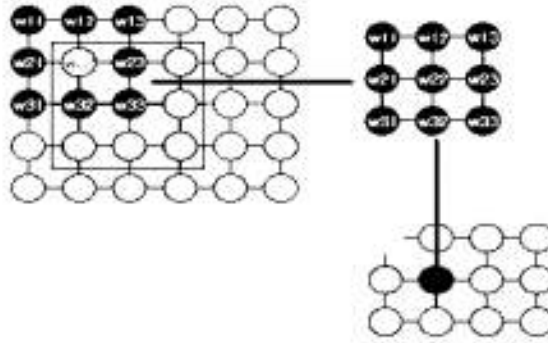


Fig 1: Pixel is multiplied with kernel filter weight producing output value (Indolia et al. 681)

From there, a pooling layer is implemented, taking the maximum or average value in a specific section of the 2D activation map, simplifying the model and reducing dimensions.

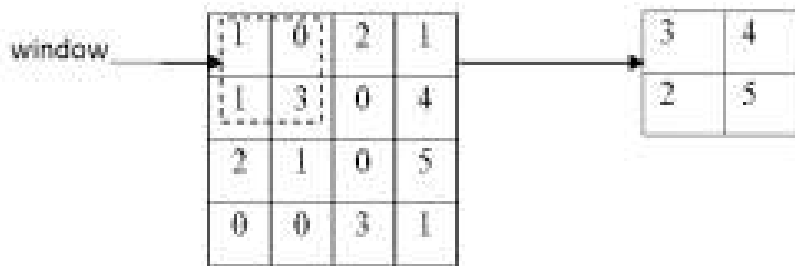


Fig 2: Dimensions are being reduced with addition of pooling layer (Indolia et al. 682)

The fully connected layer then takes the input from the pooling layer and produces a prediction for the output. The nonlinearity layer, also called the activation function, maps output values and decides the final output of the CNN (Purwono et al. 742)

We summarize the terms that are important in this paper on Table 1.

Term	Definition
Accuracy	Number of correct predictions divided by number of total predictions
Artificial Intelligence	Non-human program or model that solves complex tasks
AUC	Area Under the ROC curve represents model's ability to separate positive classes from negative classes
Batch	Set of values that are used in one training iteration

Batch size	Number of values/data samples in a batch
Binary classification	Type of classification that predicts positive class (eg: audio is deepfake) and negative class (eg: audio is real)
Dataset	Collection of raw data organized in spreadsheet or file in CSV (comma-separated values)
Dense layer	Also known as fully connected layer
Epoch	One full training pass over entire training dataset
False negative	Model mistakenly predicts the negative class.
False positive	Model mistakenly predict the positive class
Generalization	Model's ability to make correct predictions on unseen data after being trained on training data
Hyperparameter	Variables that are adjusted when tuning or training a model
Input layer	Layer of the neural network that holds the raw input data
L ₂ Regularization	Penalizes outlier weights which improves generalization of the model
Label	Answer or result of file(s) (eg: deepfake audio labeled 1, real audio labeled 0)
Loss	Answer or result of file(s) (eg: deepfake audio labeled 1, real audio labeled 0)
Machine learning	Program or system that trains model from input data to make predictions on data the model hasn't seen yet
Matplotlib	Open-source Python 2D plotting library which helps visualize aspects of machine learning
Model	Mathematical construct that takes in input and produces output
Neuron	Distinct unit within a neural network which calculates weighted sum and passes it onto the activation function
NumPy	Open-source math library in Python that provides array operations\
Overfitting	Model matches training data too closely that it can't generalize properly or make correct predictions on unseen data
Pandas	Data analysis API that is built on top of NumPy in Python.
Parameter	Weights and biases a model learns during training

Precision	True positives divided by the sum of true positives and false positives.
Preprocessing	Process and clean up data before its used to train model
Recall	True positives divided by the sum of true positives and false negatives
Regularization	Method used to reduce overfitting
Test set	Subset of dataset using for testing the model
Training	Process of determining ideal parameters in a model
Training set	Subset of dataset using for training the model
True negative	Model correctly predicts negative class
True positive	Model correctly predicts positive class
Underfitting	Model's prediction is low because it hasn't grasped the complexity of the training data
Weight	Value that model multiplies by another value

Table 1. Table of relevant machine learning terms (Machine Learning Glossary)

Literature Review

Throughout the realm of audio deepfake detection using machine learning, many researchers face challenges, especially when incorporating multiple languages in their work.

Authors Zaynab Almutairi and Hebah Elgibreen explore previous deepfake detection approaches, highlighting the need for larger and more intricate machine learning models. For example, Rodriguez-Ortega and other researchers created an H-Voice dataset, which contains voices in Spanish, English, Portuguese, French, and Tagalog. Their Logistic Regression machine learning model yields a 98 percent success rate, however the audio data needed to be manually preprocessed, requiring lots of time and effort from researchers to build the effective model. Despite the issues with the Logistic Regression model, M. Ballesteros and others utilize a CNN model called Deep4SNet, implementing it on the same H-Voice dataset. Deep4SNet displays an accuracy of 98.5 percent, slightly higher than the Logistic Regression model, but raises concerns of its own. The transformation of the audio data affects the model's performance, addressing another preprocessing issue. Similar to the Logistic Regression Model, we also manually preprocess data using a variety of techniques and methods, which requires considerable effort. However, unlike the CNN model built by M. Ballesteros and other researchers, we do not conclude that the transformation of the audio data affected the model's performance. Rather, audio data transformation is necessary to convert each audio file to the same shape, enabling our CNN model to process each file.

Wang and others developed a DNN (Deep Neural Network) model called Deep-Sonar, which achieves a 98.1 percent detection rate. Deep-Sonar was performed on the FoR (Fake OR

Real) dataset, consisting of audios in English and Chinese. However Deep-Sonar's accuracy was affected by real-world background noise, which may overestimate Deep-Sonar's accuracy. Many other researchers encounter problems causing model overfitting and poor generalization. Similarly, we also face issues regarding model overfitting, and to address this, we add an l2 regularizer. Unlike Wang and other researchers' DNN model, we do not test the extent at which real-world background noise negatively affects accuracy.

Furthermore, most datasets are trained and tested on English audios, which undervalues deepfake audios in the real world that come in a variety of dialects besides English. When experimenting with the Arabic language, researchers Amna Asif, Hamid Mukhtar, Fatimah Alqadheeb, Hafiz Farooq Ahmad and Abdulaziz Alhumam, find that deep learning approaches result in reduced accuracies due to the Arabic language's unique alphabetical pronunciation.

Authors Hasam Khalid, Shahroz Tariq, Minha Kim¹, Simon S. Woo contribute to the field by developing their own deepfake dataset, FakeAVCeleb. This dataset consists of audio and video recordings of celebrities from four ethnic backgrounds. The researchers obtain real YouTube videos from the VoxCeleb2 Dataset, and later create fake videos by generating cloned voices of the speakers from this dataset and used facial reenactment techniques to lip-sync these cloned voices with the videos. By utilizing Mel Frequency Cepstral Coefficients (MFCC) features and a variety of deepfake detection methods such as Capsule, HeadPose, and Xception, the researchers produce low Area Under the ROC curve (AUC) scores, ranging from 49 and 73 percent. Although this dataset sets a precedent for investigating audio deepfakes and is used by other researchers as part of their research, the results indicate a relatively low AUC score, highlighting the need for developing more robust models to produce more favorable results.

In this research paper, we aim to test and analyze how well a CNN model can generalize across English and Russian languages when classifying audio deepfakes. For example, if each CNN model's precision for each language is very similar, we can conclude that the CNN model may be valuable for audio deepfake detection across a multilingual context, unlike past models.

Methods

The research methodology is comprised of four distinct sections: dataset collection, data preprocessing, CNN model training, and CNN model testing. Two Convolutional Neural Network (CNN) models are built, one for Russian and one for English. In developing the CNN models, we utilized a video tutorial, which offered guidance on data preprocessing (Mulla).

Collecting Datasets

Kaggle, an online community for Data Science and Machine Learning professionals, offers tens of thousands of datasets including the Russian bona fide and spoofed speech dataset that is used. This dataset contains over 73,000 real audio clips and 128,000 fake audio clips, with each audio recording ranging from 3 to 10 seconds long (Alep).

The Fraunhofer Institute for Applied and Integrated Security AISEC is made up of researchers that explore the technology behind deepfakes and methods to enhance performance

on video and audio data. The research institute also creates a platform called DeepFake Total, which uses Artificial Intelligence to detect and analyze audio deepfakes. The platform includes datasets, such as the "In-the-Wild" dataset used in this research paper (Müller et al.). In this paper, the "In-the-Wild" dataset is referred to as the "English dataset," to avoid confusion.

The English dataset consists of 20.8 hours of real audios and 17.2 hours of fake audios of celebrities.

In order for each CNN model to be compared equally, we gave each model the same amount of input data (17.2 hours of deepfake and real audios each). We compressed the Russian dataset to 11579 deepfake and real audio clips (23158 total). We compressed the English dataset to 16508 real clips and 11816 deepfake clips. Note that the number of clips differs due to each clip varying in audio length.

Data Preprocessing

Data Preprocessing entails many key concepts before model creation: waveforms, mel spectrograms, padding, training/testing split, model training, and model testing.

Initially, we randomly selected the audio clips using the random library. After sampling the audio clips, we loaded each file using librosa, a dependency that allows for manipulation of raw audio data. Librosa converted each audio file to a NumPy array that represents amplitudes of sounds, enabling audio files to be displayed in waveforms. Later, we trimmed each audio file for efficiency purposes. We also calculated the sample rate, the number of samples taken per second of a waveform, which is implemented later on.

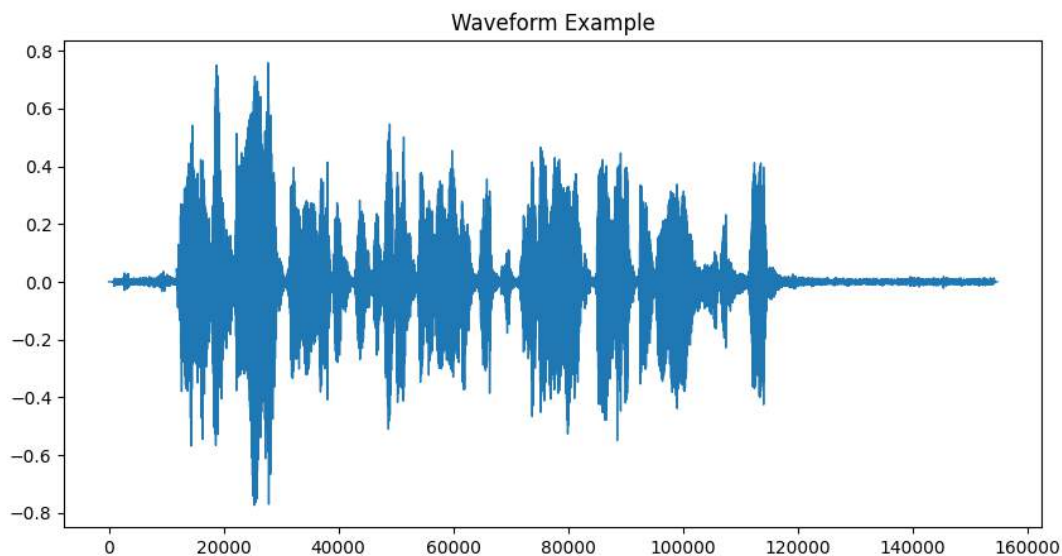


Fig 3: Waveform of 1st real Russian audio file in directory

We used the trimmed audio files and the sample rate to convert the waveforms into mel spectrograms. The mel spectrograms express unique frequencies that we hear when listening to the audio files and are critical for audio data preprocessing. We converted the amplitudes of the

NumPy arrays to decibels which are more common units when transforming audio data. The mel spectrograms accentuates the audio files' sounds and transforms them into an agreeable format for the CNN model.

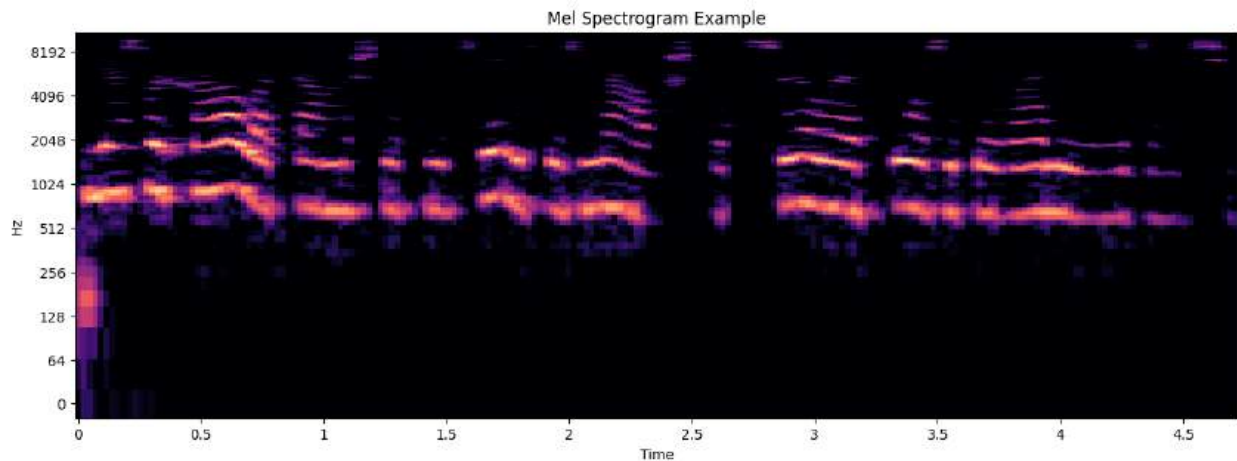


Fig 4: Mel Spectrogram of 1st real Russian audio file in directory

Now, each audio file is expressed in a mel spectrogram NumPy array, however the CNN model only works with audio files of the same shape. While the x-coordinate of each audio file's shape remained the same at 256, the y-coordinates were dissimilar. In this case, we implemented NumPy padding, where each NumPy array would have values of -80 added, resulting in each audio file having the same shape, eg: (256,450). The CNN model would disregard areas of the audio file that represent no sound or have no effect, that being the values of -80, and instead extract complex patterns on the significant areas of the audio file.

Once we thoroughly preprocessed every audio file, we split them into training (80 percent of audio files belong to training) and testing (20 percent of audio files belong to testing). If the audio file represented real audio data, its label would be 0, otherwise its label would be 1. These labels help the CNN model identify if the files are fake or real and produce statistical measurements, those being precision, recall, and loss.

Training and Testing CNN model

First, we imported a number of dependencies, allowing for the creation of the CNN model: Sequential, Conv2D, Dense, Flatten, and more. We first set the CNN model equal to Sequential, meaning that its layers are stacked sequentially. Then, we added an input layer, the first layer of the neural network, and specified the shape of the input data, eg: (256,450,1). We added Convolutional layers with relu activation, which helped the model learn complex patterns to produce an output. We also added a Flatten layer, which compresses the 2D representation of data into a single line. The dense layers with sigmoid and relu activation, connected the inputs with outputs and eventually produced the key metrics: precision, recall, and loss.

During the first time around, the CNN model produced a low initial precision compared to a high training precision. To fix this issue, we modified and tweaked several hyperparameters. Specifically, we changed the number of epochs the model ran, the batch size, adding more Convolutional and Dense layers.

The large gap between training and testing precision was particularly due to the model overfitting. This is when the CNN model is accustomed to the training data but isn't able to adapt when it views new data or testing data. This issue was solved by implementing an l2 regularizer which adds the sum of squared values of weights to the loss, penalizing larger weights and reducing overfitting. Setting the l2 regularizer to 0.002 improved the CNN model's testing precision and closely matched the precision of the training.

Results

We summarize the results in this paper on Table 2.

Language	Precision	Recall	Loss	Training and Testing time
Russian	98.09 percent	95.86 percent	0.3258	3 hrs 38 min 40 sec
English	86.89 percent	73.93 percent	2.324	8 hrs 1 min 16 sec

Table 2. Results of CNN model on English and Russian datasets

Discussion

Given the results, the CNN model performed much better (in terms of higher precision and recall and lower loss and time) when trained on the Russian dataset rather than the English dataset. The wide disparity between the CNN model's performance on the Russian dataset and the English dataset can be attributable to a variety of factors. Note that as researchers, we can't be completely sure about the particular reasons for the large discrepancy, however we can speculate about these results.

The results of the two CNN model's might be due to the differences in data quality between the Russian and English dataset. This difference can be seen under the category, "Training and Testing Time," where it took about three and a half hours for the CNN model to produce results (precision, recall, loss) unlike the CNN model operating on the English dataset, which took over eight hours to produce results. The large difference in training and testing could occur because the CNN model had a harder time processing the data for training and testing purposes.

Another reason for the large difference between training and testing times of both CNN models could be due to the number of audio clips. Both CNN models were fed 17.2 hours of real and fake audio, however the number of audio clips varied. For the Russian dataset, 11579 real and fake audio clips were fed into the CNN model. For the English dataset, 16508 real clips and

11816 fake audio clips were fed into the CNN model. The CNN model training and testing on the English dataset could have therefore taken longer to process the clips as there were more clips in this model than the model operating on the Russian dataset.

Additionally, the results could occur because some datasets (eg: Russian), perform better under certain hyperparameters and changes to the model than other datasets (eg: English). We first constructed our CNN model to operate on the Russian dataset, refining hyperparameters to yield the highest precision and recall possible (also the lowest loss possible). After our model was complete, we build the same exact model and tested it on the English dataset, which unfortunately did not produce similar results as the model operating on the Russian dataset. Implementing the l2 regularizer and changing hyperparameters may have worked better for the Russian dataset than the English. However if we first built our CNN model to operate on the English dataset, then refined our model to yield the best results, and then trained and tested the Russian model on this exact model, we may end up with different results. For example, maybe the l2 regularizer of 0.002 worked best for the Russian dataset not the English dataset, but a l2 regularizer set to 0.03 could work better for the English dataset and not the Russian dataset. Perhaps if we implemented an l2 regularizer set to 0.03, the precision and recall for the CNN model for the English dataset, could be a lot higher, and it's loss could be lower.

The difference between the English dataset and Russian dataset's performance may also be attributed to the diversity of the speakers in each dataset. The English dataset contains hours of English audio from 58 celebrities and politicians, each with a unique diction and style of speech due to factors such their upbringing. For example, Arnold Schwarzenegger, a former pro-body builder from Austria, has a different perception of the English language and accent compared to Donald Trump, the former president of the United States who grew up in New York. These disparities in speaking style can add to the difficulty of classifying real and fake audio. On the other hand, the Russian dataset does not specify whether the audio recordings are from native Russian speakers or others. Therefore, it is not entirely clear if the diversity of the speakers accounts for the differences in the models' results for each language.

When we built the CNN models, we also faced lots of limitations and challenges. Initially, we planned to test how a CNN model would operate on English, Chinese, and Russian datasets. However, when inputting 17.2 hours of Chinese audio data into the CNN model for the Chinese dataset, the kernel often crashed. We troubleshooted this problem by reducing the number of hours of input data significantly, however the kernel continued to crash. This issue could largely be due to the audio quality of the Chinese dataset or because we fed a larger number of clips into the CNN model for the Chinese dataset than that of the other models.

Another limitation and challenge is the lack of computational power, since we didn't have access to supercomputers when running our models. With supercomputers or highly advanced software, the time for training and testing the model would significantly decrease and our model's performance on the Chinese and English datasets may increase due to the technology's ability to better handle larger amounts of data.

Conclusion

We present a machine learning approach to test whether a Convolutional Neural Network (CNN) model can generalize across English and Russian languages when classifying deepfake audios. We identify a variety of data preprocessing concepts and techniques to clean and prepare our data for training and testing the CNN model, such as data trimming, waveforms, and mel spectrograms. We add multiple layers to create our CNN model, update hyperparameters, and include an l2 regularizer, allowing the model to achieve greater results. We compare the performance of the models for the English and Russian dataset. Although our CNN model faces challenges generalizing across the English and Russian datasets, we believe that our research can set a precedent for future research in identifying effective methods to differentiate real and deepfake audios across multiple languages. We also aim for future research to address the limitations and challenges we encounter during our research process.

Acknowledgment

I would first like to thank my Research Mentor, Hayyu Imanda, for her invaluable guidance and feedback throughout this research process. I am also grateful for the Lumiere Research Scholar Program, for providing this exceptional opportunity. I would like to thank my Program Manager, Yamini Joshi, for her support and kindness throughout this journey. Lastly, I would like to thank my friends and family for their encouragement.

Works Cited

- Alep. "Pyara: Russian Bona Fide and Spoofed Speech." Kaggle, 27 June 2023, www.kaggle.com/datasets/alep079/pyara.
- Almutairi, Zaynab, and Hebah Elgibreen. "A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions." MDPI, Multidisciplinary Digital Publishing Institute, 4 May 2022, www.mdpi.com/1999-4893/15/5/155.
- Alzubaidi, Laith, et al. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions - Journal of Big Data." SpringerOpen, Springer International Publishing, 31 Mar. 2021, journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8.
- Asif, Amna, et al. "An Approach for Pronunciation Classification of Classical Arabic Phonemes Using Deep Learning." MDPI, Multidisciplinary Digital Publishing Institute, 27 Dec. 2021, www.mdpi.com/2076-3417/12/1/238.
- Ballesteros, Dora, et al. "A Dataset of Histograms of Original and Fake Voice Recordings (H-Voice)." Data in Brief, Elsevier, 26 Feb. 2020, www.sciencedirect.com/science/article/pii/S2352340920302250.
- Ballesteros, Dora, et al. "Deep4SNet: Deep Learning for Fake Speech Classification." Expert Systems with Applications, Pergamon, 29 June 2021, www.sciencedirect.com/science/article/abs/pii/S0957417421008770.
- "Creative Commons." Deed - Attribution-NonCommercial-ShareAlike 4.0 International - Creative Commons, creativecommons.org/licenses/by-nc-sa/4.0/.
- "Deepfakes - Fraunhofer AISEC." Fraunhofer Institute for Applied and Integrated Security, 27 Aug. 2024, www.aisec.fraunhofer.de/en/about-us/scientific_excellence/Deepfakes.html.
- George, Shaji, and Hovan George. "Deepfakes: The Evolution of Hyper Realistic Media Manipulation." ResearchGate, www.researchgate.net/publication/375697835_Deepfakes_The_Evolution_of_Hyper_realistic_Media_Manipulation.
- Indolia, Sakshi, et al. "Conceptual Understanding of Convolutional Neural Network- a Deep Learning Approach." Procedia Computer Science, Elsevier, 8 June 2018, www.sciencedirect.com/science/article/pii/S1877050918308019.
- Khalid, Hasam, et al. "Fakeavceleb: A Novel Audio-Video Multimodal Deepfake Dataset." arXiv.Org, 1 Mar. 2022, arxiv.org/abs/2108.05080.
- "Machine Learning Glossary | Google for Developers." Google, Google, developers.google.com/machine-learning/glossary.
- Mahmud, Bahar, and Afsana Sharmin. "Deep Insights of Deepfake Technology : A Review." ResearchGate, www.researchgate.net/publication/351300442_Deep_Insights_of_Deepfake_Technology_A_Review.
- Mulla, Rob. "Audio Data Processing in Python." YouTube, 24 Feb. 2022, www.youtube.com/watch?v=ZqpSb5p1xQo.
- Müller, Nicolas M., et al. "Does Audio Deepfake Detection Generalize?" Interspeech, 2022.
- Purwono, Puwono, et al. "Understanding of Convolutional Neural Network (CNN): A Review." ResearchGate, www.researchgate.net/publication/367157330_Understanding_of_Convolutional_Neural_Network_CNN_A_Review.
- Wang, Run, et al. "Deepsonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices." arXiv.Org, 15 Aug. 2020, arxiv.org/abs/2005.13770.
- "Your Machine Learning and Data Science Community." Kaggle, www.kaggle.com/.

Using Prudence to Predict the International Electric Vehicle Market By Ryon Das

Abstract:

In this paper, I study the current international market revolving around electric vehicles (EVs), with an emphasis on the countries of the United States and China. I use a Nash Equilibrium model within a game theoretic framework to assess the outcomes of possible actions taken by these two leading countries. Specifically, I structure this scenario as a multistage game, in which I use a variable I have defined as Prudence to determine the Nash Equilibria of various tariff scenarios by varying the foresightedness levels of the countries towards global needs, rather than short-term economic squabbles. I then provide policy recommendations for a global entity such as the UN to carry out in order to secure international emissions targets best. The outcomes of the model show that there exists a wide range of values for prudence in which the actions of the US and China lead to problematic outcomes for the world. As a result, this paper demonstrates the importance of global action in order to prevent catastrophic outcomes in the realm of the global fight against climate change, specifically regarding Electric Vehicles.

Introduction:

Electric vehicles (EVs), and the transition away from traditional gas vehicles, has the potential to prevent 800 million tons of CO₂ every year by 2040. They are a revolutionary product, one which can help prevent the planet from reaching the terrifying effects of Climate Change forecasted. But on a geopolitical level, whoever can control the creation, proliferation, and innovation of this product is the country who can lead the world into a future of environmentally driven policy and technology, which might be the challenge of our time.

EVs are not a newfound invention; they were well-used in the late 1800s and early 1900s. But with the creation of cost-efficient gas automobiles, such as the Model T of Ford, they took a backseat for the next century. But at the beginning of the new millennium, with projects such as Tesla Motors or the Prius model of Toyota, an electric-gas hybrid vehicle, electric vehicles were back in vogue again (Matulka). And it was of the utmost importance that they were, as oil shortages in the eighties, along with the increasing issue of climate change opened the public up to trying out new things.

However, while EVs have increased in the US the last two decades, the US does not remain at the forefront of the market anymore. Over the last five years, Tesla's dominance has waned, both domestically and abroad, as American customers continue to struggle to make the switch away from traditional vehicles and as prices internationally remain high. In fact, in the first quarter of 2024, Tesla has seen its sales down 20% and its share price decrease by 25% compared to the same period in 2023 (Oxford).

Rather, the new players in the market are Chinese EV companies such as BYD and GAC Aion, as well as other Chinese companies like Xiaomi, originally a phone company, now launching Electric Vehicle models. Chinese EV companies now make up 60% of the worldwide

sales of EVs, up from 0.1% at the beginning of the 2010s. Chinese companies now hold three out of the top five global market share spots for EVs (Oxford).

The reason for all of this is, as previously mentioned, the price of the Chinese vehicles. While they are not currently sold in the United States, in other locations, they are much more affordable to the average consumer. For instance, in Mexico, a Dolphin Mini from BYD would cost \$21,000, while a Chevrolet Bolt would cost \$27,000 (Oxford).

This is a bitter pill for the United States Government to swallow. As a result, two successive presidents have incorporated import tariffs on China. First, President Trump increased the tariff rate on Chinese vehicles from the standard 2.5% to 25%. And in the spring of 2024, President Biden doubled down on this policy, increasing the tariff rate to 100% (Tankersly). To put this into perspective, let's say a car imported from China that a Chinese company is selling is valued at \$100. In a perfect laissez-faire model, the consumer would be able to buy it at the \$100 mark. But under Trump-era policies, that car would now be available to be bought at \$125, with the government pocketing the rest. And the new policies of Biden would make this car be sold at no less than \$200, with the full 100% tariff.

This becomes interesting when viewing the price of American EVs. Teslas are now the cheaper option compared to the Chinese EVs, despite the lower original market price of the Chinese EVs. As a result, more consumers will buy American vehicles, and the goal of policies like this is to establish a firmer American market and more competitive, both quality and pricewise, American EV companies.

Unfortunately, this is not likely to go over very well with policymakers in Beijing. As a result of the actions of the United States, their ability to spread their own EVs is largely diminished. Moreover, the US sanctions increase the possibility of other polities, such as the EU or Japan (Lynch et al.), imposing their own tariffs. China thus has a difficult choice on whether to respond or not to these sanctions. And depending on China's choice, the United States will also have a decision to make on how to further the situation, which I will discuss more in the following sections of the paper. As a result, there are many scenarios which can occur from the situation in the status quo.

However, it is clear that for the world, the best scenario involves a free and open market for EVs which allows the most amount of innovation and cost-reduction of vehicles, allowing both greater use as well as better vehicles. This would increase the likelihood of meeting necessary emissions targets, as well as minimize the additional carbon footprint caused by gas-powered vehicles. The question my paper thus attempts to ask is how can the UN, or an undefined international body of countries, attempt to keep the market of Electric Vehicles as open as possible, taking into account the variable of how much the US and China are interested in the future?

To answer this, I have developed a sequential and responsive game theoretic model which revolves around a three-form output structure, which incorporates both short term economic desires as well as long-term climate change goals. I run multiple analyses of my model under different weightings of the variable I term Prudence, which I use to represent the value that each

country places on the future. Here, P_U and P_C denote the prudence of the US and China, respectively. Under each set of conditions, I solve for the Nash Equilibria, the outcome that results when each player plays the best move given the move of the other player. In particular, as I assign values to describe the outputs of each decision-making tree, the Nash Equilibrium for this model would be where the US and China both maximize their numerical output given what they know about the other's decision making.

While Nash Equilibria are a common tool to utilize in almost all game theoretical models, in this model specifically I will rely on solving for the Subgame Perfect Nash Equilibria or SPNE. In sequential games like this one, to analyze the most likely decision, I will find the most likely play of the US given either of China's actions. Then, if China knows that the US will do one of two possibilities, China will do the action which leads to the US decision which China prefers more. I will find the SPNE of the model under different sets of prudence constants.

Through the results of my study, I have found that at Prudence values of $P_U \geq 0.16$ and $P_C \geq 0.33$, the world's output of EVs is maximized. However, these numbers illustrate that both countries must place high priority on future-oriented thinking, which is far from guaranteed. As a result, the United Nations should attempt to create a form of "positive reinforcement," which essentially means to increase the outputs for the US and China in the situations which are beneficial for international climate change efforts. Possible policy recommendations include instituting leadership positions in committees or reducing contributions in UN dues for the US and China in an effort to increase the numerical benefit for the countries in the situations where they withdraw sanctions. Therefore, these recommendations can act as incentives to shift the results of our experiment such that at even lower P_U and P_C values, the countries still choose the world-optimal scenario.

Methods:

I: Set-up

In this game, for both China and the United States, there are short-term consequences, such as jobs created or additional sales that tend to favor one side, as well as long-term consequences, such as preventing climate change emissions that are net-beneficial for the world. As a result, every outcome can be represented in the payoff form (S_U, S_C, L) , where S_U is the short-term outcome for the US, S_C is the short-term outcome for China, and L is the long-term outcome for the world. I will refer to this three outcome model as the primitive payoff form.

However, a model with three payoff parts which only corresponds to two actors is challenging to properly analyze. As a result, I use the variables P_U and P_C , which represent the prudence of the United States and China, respectively. P can take on any real value between 0 and 1, and represents the weighting of the actor's preference towards long-term interests rather than a focus on short-term interests. Thus, the payoff for the US, which I will characterize as U , can be determined from the primitive payoff form with the expression:

$$S_U \cdot (1 - P_U) + L \cdot P_U$$

whereas the payoff form for China, or C can be represented in the form:

$$S_C \cdot (1 - P_C) + L \cdot P_C$$

We refer to the payoff form of (U, C) as the final payoff form. While we use the same model and the same primitive outcomes for every example, the final payoff forms vary from instance to instance as the Prudence constant of the actors is shifted.

II: Initial Game

As of this writing (July 2024), the United States has already made the first move, increasing the Chinese EV tariff to 100%. As a result, this model begins with China's choice on whether to impose retaliatory sanctions or not. While this might seem like a clearcut decision, China understands that the United States is uniquely in an unstable position, with the forthcoming 2024 election, and as a result, the incumbent party needs to appear tough on China. As a result, the PRC understands that if they get into a tariff battle, there is a strong risk of delving into a trade war. This creates the first stage of the game, as China is forced to weigh its actions to determine the best set of outcomes.

As a result, China has two clear choices: A retaliatory tariff or simply doing nothing. This choice is slightly oversimplified, but the overall decision is a true one: namely, whether China wants to get embroiled in an escalatory situation or whether they prefer to lose money instead.

Let us assume that China does decide to respond with a retaliatory tariff. That would mean the US has three responses to this: the US could raise the stakes even further and get into a full-scale trade battle; the US could do nothing but maintain the original tariff; and finally, the US could lift the sanctions to deescalate the situation.

Alternatively, if China decides to risk going into a trade war and instead takes the loss, the US also has two potential responses. They can increase the tariff further and make it even broader, or they can simply let it be.

It is also important to note that while the US has three responses if China retaliates to the tariff, there are only two logical possibilities if China does not. After all, why would the US back down their tariffs if China does not respond in the first place?

With these sets of choices, the diagram which represents this game is as follows.

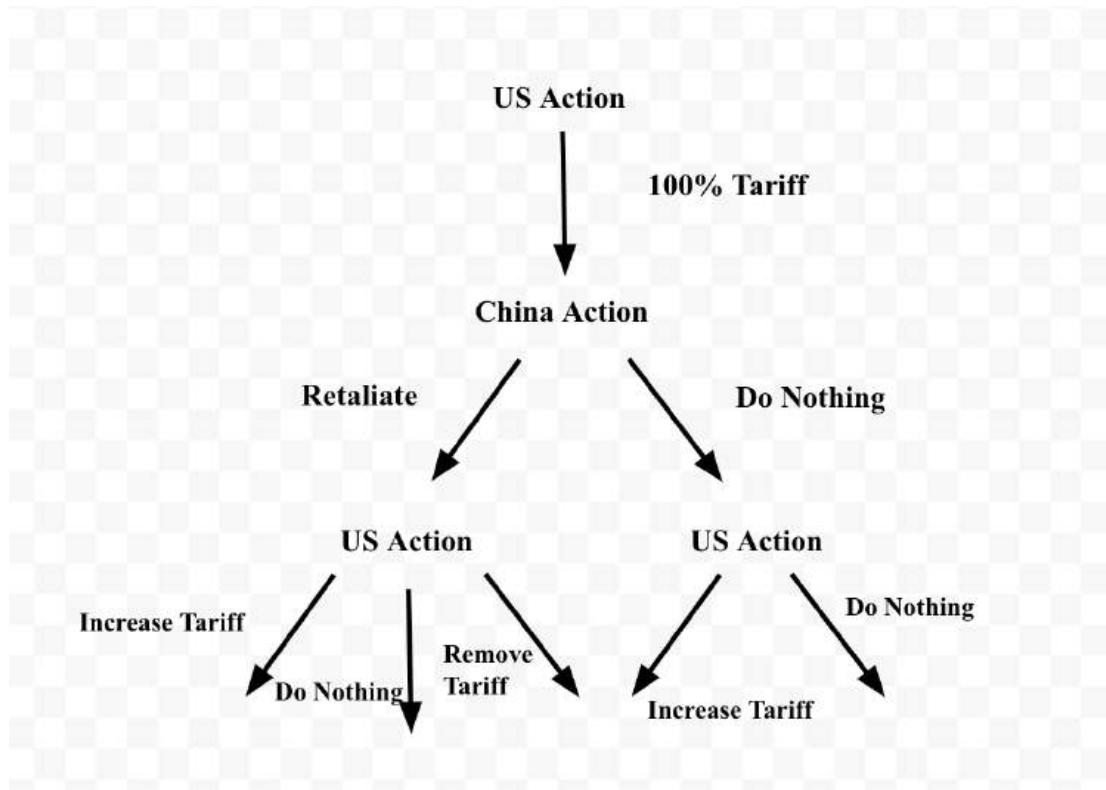


Figure 1: Initial Model Figure

But this model is incomplete at this point. Now, the values in the primitive payoff form need to be determined. Let us examine the outcomes of all five possibilities, determining somewhat arbitrary values for each set. While there is no mathematical reason to state that the long-term coefficient for the both countries' do-nothing scenario is 50, the more important factor at play is the relativity. I will be attempting to assign values such that each output value for each possibility is approximately relative to the others, which means that I will have to make some arbitrary assignments.

Let us start with the possibility that both China and the United States do nothing. This would likely have a neutral effect on the US and a slightly negative economic/short-term effect on China. This would result from the already-imposed tariffs pricing Chinese electric vehicles out of certain markets, but certainly not crippling them. However, the long-term coefficient would have a large positive value. This outcome represents the freest market for EVs in the future, pushing innovation of better EVs and allowing the largest number of people to gain access to EVs (Wagner and Walsh). And because EVs on net in their present state have lower vehicular CO₂ emissions by 43% compared to traditional alternatives (Hausfather), the proliferation of EVs would see a great benefit to the world in terms of reducing the effects of climate change. Thus, if we were to propose an abstract payoff set, which again is arbitrary and attempts to show the relative impacts of the choice towards each player, the values would be (0, -5, 50), in the primitive model structure.

The next possibility that exists is where China does nothing, and the US, sensing a possibility, decides to increase tariffs further. This would hurt China further, while helping the United States standing in the Electric Vehicle Market. However, the global long-term coefficient would also decrease relative to the former scenario as it becomes harder to own Electric Vehicles due to increased pricing. As a result, the values would be something like (10, -20, 0).

Now, we must consider the possibility that China actually retaliates. If the United States attempts to counter, this will harm both countries economically due to the trade war, along with having a hugely detrimental impact on the global long-term coefficient, with the most closed off market and risks to stability further hampering efforts to combat climate change. Thus, I have assigned the payoff matrix being (-10, -10, -50).

A different possibility is the chance that the US does nothing in response to China's retaliatory tariffs. Again, neither country would benefit in a situation like this, although the global coefficient would not be as dramatically reduced. The payoff matrix is (-10, -5, -20).

The final situation for which we must determine payoffs is the one most in China's favor, where they retaliate and the United States responds by backing down and removing their tariff in response. This would also be beneficial for the global coefficient, as it would see a somewhat free market, similar to the first possibility outlined. However, it would also be a stunning loss for the United States' short-term coefficient. The matrix would look like (-15, 10, 20).

As a result, we have the following model for the initial stage.

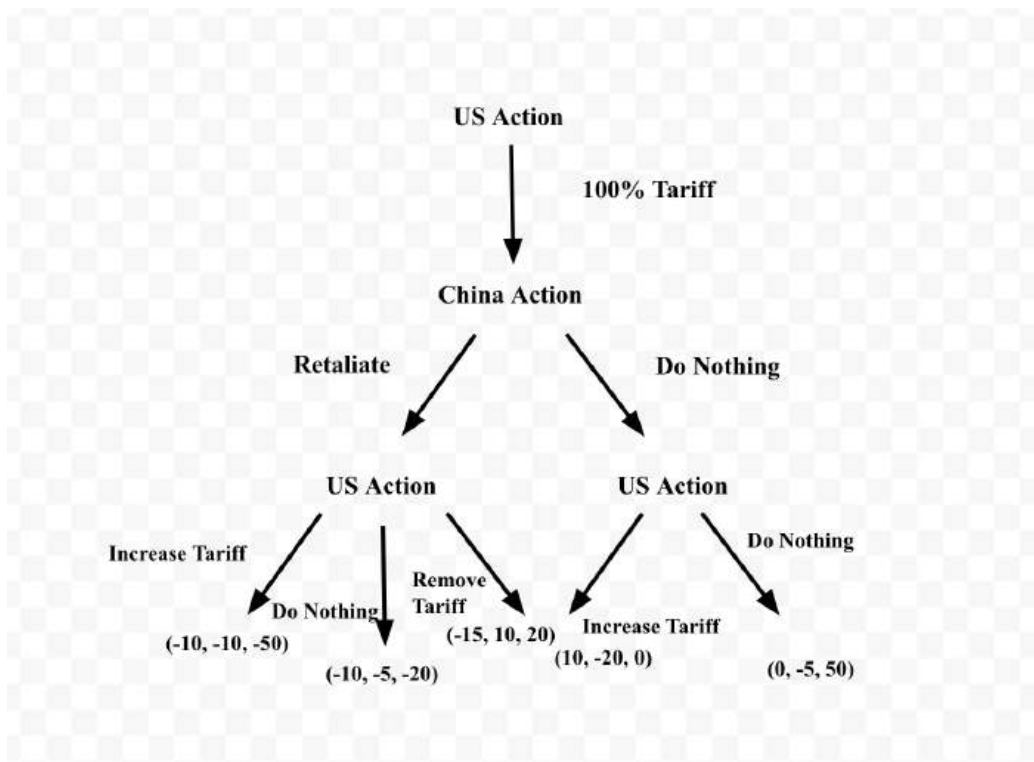


Figure 2: Model Figure with Values

Alternatively, we can look at the possibilities in a table form.

(S_U, S_C, L)	China responds	China does nothing
US increases Tariff	(-10, -10, -50)	(10, -20, 0)
US does nothing	(-10, -5, -20)	(0, -5, 50)
US removes Tariff	(-15, 10, 20)	

Table 1: Model Figure with values in table form

III: Initial Game Prudence Analysis

Let's begin by testing some extreme cases by using the formulae we have previously defined:

$$((S_U \cdot (1 - P_U)) + L \cdot P_U)$$

$$((S_C \cdot (1 - P_C)) + L \cdot P_C)$$

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -10)	(10, -20)
US does nothing	(-10, -5)	(0, -5)
US removes Tariff	(-15, 10)	

Table 2: If $P_U = 0, P_C = 0$ (We assume both countries are completely shortsighted)

As this is a sequential game, we should start from the end, and examine what the United States does. If China responds, the US's best course of action is to increase the tariff or do nothing. And if China does not respond, the US's best course of action is to increase the tariff. As a result, China knows that there are three scenarios possible, and the worst scenario for China would be to do nothing as this yields an output of -20. As a result, China will always respond, which means that the subgame perfect Nash Equilibria of this game are China responding and the US increasing the tariff or doing nothing. But for the sake of this simulation, we assume that the US would prefer what China prefers less, meaning that the US increasing the tariff is the most likely scenario.

(U, C)	China responds	China does nothing
US increases Tariff	(-50, -50)	(0, 0)
US does nothing	(-20, -20)	(50, 50)
US removes Tariff	(20, 20)	

Table 3: If $P_U = 1, P_C = 1$ (We assume both countries are completely farsighted)

By the same process, solving backwards, we see that the US prefers to remove the tariff if China responds, and does nothing if China does nothing. The best scenario for China would be to do nothing: (50, 50).

(U, C)	China responds	China does nothing
US increases Tariff	(-50, -10)	(0, -20)
US does nothing	(-20, -5)	(50, -5)
US removes Tariff	(20, 10)	

Table 4: If $P_U = 1$, $P_C = 0$ (We assume the US is completely farsighted and China is completely shortsighted)

There are only two possibilities (20,10) and (50, -5), by the same process for the US. China looks and sees the best for them is (20, 10).

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -50)	(10, 0)
US does nothing	(-10, -20)	(0, 50)
US removes Tariff	(-15, 20)	

Table 5: If $P_U = 0$, $P_C = 1$ (We assume the US is completely shortsighted and China is completely farsighted)

Now, we can proceed ahead and start checking intermediate cases, again using the same process to find the Subgame Perfect Nash Equilibriums.

(U, C)	China responds	China does nothing
US increases Tariff	(-18, -18)	(8, -16)
US does nothing	(-12, -8)	(10, 6)
US removes Tariff	(-8, 12)	

Table 6: If $P_U = 0.2$, $P_C = 0.2$

(U, C)	China responds	China does nothing
US increases Tariff	(-26, -26)	(6, -12)
US does nothing	(-14, -11)	(20, 17)
US removes Tariff	(-1, 14)	

Table 7: If $P_U = 0.4, P_C = 0.4$

(U, C)	China responds	China does nothing
US increases Tariff	(-34, -34)	(4, -8)
US does nothing	(-16, -14)	(30, 28)
US removes Tariff	(6, 16)	

Table 8: If $P_U = 0.6, P_C = 0.6$

(U, C)	China responds	China does nothing
US increases Tariff	(-42, -42)	(2, -4)
US does nothing	(-18, -17)	(40, 39)
US removes Tariff	(13, 18)	

Table 9: If $P_U = 0.8, P_C = 0.8$

We can clearly see at higher prudence levels that the Nash Equilibrium results in the US doing nothing and China doing nothing. Moreover, this change occurs between the P levels of 0.2 and 0.4. If we view this from the world's perspective, this is clearly the optimal outcome for global emissions rates; as a result, I will refer to it as optimal from now on. Consequently, let us examine some asymmetric intermediate cases, where the US and China have different P levels, to further make observations.

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -10)	(10, -20)
US does nothing	(-10, -5)	(0, -5)
US removes Tariff	(-15, 10)	

Table 10: If $P_U = 0, P_C = 0$

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -18)	(10, -16)
US does nothing	(-10, -8)	(0, 6)
US removes Tariff	(-15, 12)	

Table 11: If $P_U = 0$, $P_C = 0.2$

This condition is unique. Off the bat, the US prefers either (-10, -18), (-10, -8) if China responds, and (10, -16) if China does nothing. However, China does not know whether the US will select (-10, -18) or (-10, -8), so its best strategy is to play (10, -16).

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -26)	(10, -12)
US does nothing	(-10, -11)	(0, 17)
US removes Tariff	(-15, 14)	

Table 12: If $P_U = 0$, $P_C = 0.4$

The same process holds true for this model as well.

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -34)	(10, -8)
US does nothing	(-10, -14)	(0, 28)
US removes Tariff	(-15, 16)	

Table 13: If $P_U = 0$, $P_C = 0.6$

(U, C)	China responds	China does nothing
US increases Tariff	(-10, -42)	(10,)
US does nothing	(-10, -17)	(0,)
US removes Tariff	(-15, 18)	

Table 14: If $P_U = 0$, $P_C = 0.8$

(U, C)	China responds	China does nothing
US increases Tariff	(-18, -10)	(8, -20)
US does nothing	(-12, -5)	(10, -5)
US removes Tariff	(-8, 10)	

Table 15: If $P_U = 0.2$, $P_C = 0$

(U, C)	China responds	China does nothing
US increases Tariff	(-18, -18)	(8, -16)
US does nothing	(-12, -8)	(10, 6)
US removes Tariff	(-8, 12)	

Table 16: If $P_U = 0.2$, $P_C = 0.2$

(U, C)	China responds	China does nothing
US increases Tariff	(-18, -26)	(8, -12)
US does nothing	(-12, -11)	(10, 17)
US removes Tariff	(-8, 14)	

Table 17: If $P_U = 0.2$, $P_C = 0.4$

(U, C)	China responds	China does nothing
US increases Tariff	(-18, -34)	(8, -8)
US does nothing	(-12, -14)	(10, 28)
US removes Tariff	(-8, 16)	

Table 18: If $P_U = 0.2$, $P_C = 0.6$

There is no need to continue checking scenarios; we can see that the prudence-level checkmark for China to change its decision making towards optimal is again between 0.2 and 0.4. Now, we can move on to the 0.4 threshold.

(U, C)	China responds	China does nothing
US increases Tariff	(-26, -10)	(6, -20)
US does nothing	(-14, -5)	(20, -5)
US removes Tariff	(-1, 10)	

Table 19: If $P_U = 0.4$, $P_C = 0$

(U, C)	China responds	China does nothing
US increases Tariff	(-26, -18)	(6, -16)
US does nothing	(-14, -8)	(20, 6)
US removes Tariff	(-1, 12)	

Table 20: If $P_U = 0.4$, $P_C = 0.2$

(U, C)	China responds	China does nothing
US increases Tariff	(-26, -26)	(6, -12)
US does nothing	(-14, -11)	(20, 17)
US removes Tariff	(-1, 14)	

Table 21: If $P_U = 0.4$, $P_C = 0.4$

Again, there is no need to continue checking; we can clearly see that between 0.2 and 0.4, China begins to prefer the optimal scenario. A pattern has begun to emerge, but we verify it with an additional set of cases.

(U, C)	China responds	China does nothing
US increases Tariff	(-34, -10)	(4, -20)
US does nothing	(-16, -5)	(30, -5)
US removes Tariff	(-6, 10)	

Table 22: If $P_U = 0.6$, $P_C = 0$

(U, C)	China responds	China does nothing
US increases Tariff	(-34, -18)	(4, -16)
US does nothing	(-16, -8)	(30, 6)
US removes Tariff	(-6, 12)	

Table 23: If $P_U = 0.6$, $P_C = 0.2$

(U, C)	China responds	China does nothing
US increases Tariff	(-34, -26)	(4, -12)
US does nothing	(-16, -11)	(30, 17)
US removes Tariff	(-6, 14)	

Table 24: If $P_U = 0.6$, $P_C = 0.4$

We see a pattern forming that $P_U \in \{0.2, 0.4, 0.6\dots\}$, if $P_C = 0$ or 0.2 then we see the China responding and US removing scenario, whereas at $P_C > 0.4$, the optimal scenario occurs. The same holds true for $P_U = 0.8$.

We can thus create the following table to demonstrate our findings, where the values of the US are demonstrated on the left column and the values of China are demonstrated on the top row.

Moreover, we can represent the outcomes in terms of global benefit through a color gradient in the following manner, where the greener colors are more beneficial for the world and the redder colors are the least beneficial for the world. The map gives the colors that correspond to outcomes in the following table.

(S_U, S_C, L)	China responds	China does nothing
US increases Tariff		
US does nothing		
US removes Tariff		

Table 25: Color-coordinated map of possibilities

	$P_C = 0$	$P_C = 0.2$	$P_C = 0.4$	$P_C = 0.6$	$P_C = 0.8$	$P_C = 1$
$P_U = 0$						
$P_U = 0.2$						
$P_U = 0.4$						
$P_U = 0.6$						
$P_U = 0.8$						
$P_U = 1$						

Table 26: Nash Equilibrium at all possibilities

This table of tables confirms the trend we have been witnessing. As long as the US has a P_U value of 0.2 or more, the P_C value determines whether we will see the light green scenario in the table or the optimal scenario. Moreover, we see that there is some threshold of P_C between 0.2 and 0.4, where, before the threshold is passed, the former scenario plays out, and after the threshold is passed, the latter scenario plays out.

Further calculations isolate the true value for this threshold where optimal is reached. With the equation

$$10(1-P_C) + 20P_C = -5(1-P_C) + 50P_C$$

When we attempt to solve this, we find that the threshold (the one we know to be between 0.2 and 0.4) is exactly $\frac{1}{3}$.

We can also find the P_U threshold value, all though it is not as abundantly clear as the P_C value. However, we can see that if P_C is sufficiently high, that there is a P_U value between 0 and 0.2 where the Nash Equilibrium goes from being the yellow scenario or the optimal scenario. We can again set an equation.

$$10(1-P_U) = 50P_U$$

Solving this, we find that the threshold for P_U is $\frac{1}{6}$.

As a result, we have solved the model. The United Nations should know that only if $P_U > 0.1\bar{6}$ and $P_C > 0.3\bar{3}$ hold true does the optimal scenario occur.

Conclusion and Policy Recommendation:

We have solved all quintile possibilities. We have determined the levels at which the United States and China have preference angling towards the outcome which is best for the

world. Now, it is critical to reflect on how it would be possible, as an international body, to angle the US and China into this scenario.

I have thought of two possibilities. The first would be to guarantee that both the United States and China have P-values above their thresholds of $P_U > 0.1\bar{6}$ and $P_C > 0.3\bar{3}$. Unfortunately, this would be rather hard to do. International bodies are notoriously inefficient at directly being able to influence countries in this manner (Crisis Group). Moreover, this process would raise further questions of this model in particular. The values are not perfectly relative to the actual outcomes in the real world, nor do they strive to be. As a result, this path would not be very feasible, given that this paper serves more abstract purposes when creating the model than perfectly specific ones.

The other possibility would be for the United Nations to attempt to affect the values associated with the game. For instance, finding a way to decrease the short-term coefficients further in the suboptimal scenarios is one potential path to effectively shift the game. This could look like sanctions or tariffs to the United States or China in scenarios where either or both decide to further continue this trade battle. Unfortunately however, this would also be unlikely to occur in the real world, given the United States and China's power and prestige on the five permanent members of the Security Council, and their ability to veto any measures they find hard to stomach (Security Council Report). This form of compulsion, of attempting to decrease certain values, would likely fall flat on its head.

As a result, the only possibility left standing is what can be described as a type of positive reinforcement. Increasing the short-term outputs of the US and China in the optimal conditions, is the most feasible of the scenarios. An idea to consider includes incentivizing cooperation by offering incentives for the US and China if they withdraw sanctions. For example, if the UN were to offer leadership roles on committees or initiatives, or by decreasing the contributions that the countries have to provide to the UN (Council on Foreign Relations), in exchange for the countries removing sanctions on the UN, this might promote the countries to follow the optimal scenario, which would undoubtedly be the best for emissions efforts.

Works Cited

- Council on Foreign Relations. "Funding the United Nations." Council on Foreign Relations, Council on Foreign Relations, 8 June 2020, www.cfr.org/article/funding-united-nations-what-impact-do-us-contributions-have-un-agencies-and-programs.
- Crisis Group. "Ten Challenges for the UN in 2023-2024." www.crisisgroup.org, 14 Sep. 2023, www.crisisgroup.org/global/sb11-ten-challenges-un-2023-2024.
- Hausfather, Zeke. "Factcheck: How Electric Vehicles Help to Tackle Climate Change." Carbon Brief, 13 May 2019, www.carbonbrief.org/factcheck-how-electric-vehicles-help-to-tackle-climate-change/.
- Lynch, Suzanne, et al. "The U.S. Wants Europe to Stand up to China. Europe Says: Not so Fast." POLITICO, 8 Mar. 2023, www.politico.com/news/2023/03/08/us-europe-china-00086204.
- Matulka, Rebecca. "The History of the Electric Car." Energy.gov, Department of Energy, 15 Sept. 2014, www.energy.gov/articles/history-electric-car.
- Oxford, Dwayne. "Are Chinese Electric Vehicles Taking over the World?" Al Jazeera, 2024, www.aljazeera.com/economy/2024/4/20/are-chinese-evs-taking-over-the-car-market.
- Security Council Report. "The Veto : UN Security Council Working Methods : Security Council Report." Securitycouncilreport.org, 13 Feb. 2024, www.securitycouncilreport.org/un-security-council-working-methods/the-veto.php.
- Tankersley, Jim. "Biden's China Tariffs Are the End of an Era for Cheap Chinese Goods." The New York Times, 18 May 2024, www.nytimes.com/2024/05/18/business/biden-china-tariffs.html#:~:text=A%20wave%20of%20imports%20from.
- Wagner, Gernot, and Conor Walsh. "Opinion | Don't Slam the Door on Inexpensive Chinese Electric Vehicles." The New York Times, 15 May 2024, www.nytimes.com/2024/05/15/opinion/biden-china-trade-electric-vehicles-solar-panels-tariffs.html.

Expanding Affordable Housing in Northern Virginia: Addressing Gaps in “Live, Work, Play” Communities By Sophie Zhang

Abstract

“Live, work, play” communities are designed to combine residential, commercial, and recreational spaces into one neighborhood. However, housing prices have significantly increased due to rising demand in these areas, making homes unaffordable for many workers. Without adequate strategies, residents of Fairfax County could be left homeless and inequality could increase. This paper explores current affordable housing efforts in two of the largest “live, work, play” neighborhoods in Fairfax County, Tysons Corner and the Mosaic District. Through these case studies, the paper offers specific strategies and recommendations that have worked and should be implemented in future developments, such as public-private funding partnerships and the importance of location. The policy suggestions aim to expand Fairfax County’s affordable housing stock and support the county’s commitment to long-term affordability.

Live, Work, Play Communities

The term “live, work, play” communities was coined to describe districts or urban hubs where residential, commercial, and recreational areas are integrated, allowing residents to have practically their whole world in the same area (McGrath). These communities aim to reduce the long commutes for employees and create a convenient space for both living and working. However, the increasing desire for these residential areas have caused housing prices to skyrocket, leaving employees with no other choice but to find housing elsewhere, eroding the very purpose of these communities. In fact, within a seven month period at the beginning of 2024, rents in Fairfax County have risen 9.5% (“Fairfax, VA Rental Market”). Without new policies or solutions, these communities can become “exclusive” and out of reach for the essential employees who sustain them. Thus, it is crucial that Northern Virginia implements policies and strategies to increase affordable housing within “live, work, play” communities, ensuring that all residents, regardless of income, have access to these vibrant spaces. The following case studies will examine two of the largest “live, work, play” communities in Northern Virginia and evaluate their current affordable housing initiatives.

For the purposes of this paper, affordable housing is defined as housing that costs no more than 30% of a household’s income (§ 15.2-2201). This standard is essential to ensure families have enough resources to provide for other necessities such as food, clothing, and transportation. Second, in Fairfax County, VA, the low income limit for a family of four is \$97,800 annually. This income group plays an important role in our economy as they hold a significant portion in key work sectors such as the service and retail industry.

Case Studies

Tysons Corner

Carmen Romero, head of the Arlington Partnership for Affordable Housing, notes that despite having over 100,000 workers in the Tysons, Virginia area, there have been no major affordable housing developments (Kimura). This is all about to change as The Exchange, projected to open in 2027 by Spring Hill Metro Station, will be Tysons Corner's first 100% affordable housing development. This 516-unit building will only be eligible for people earning between 30-70% of the area median income in Tysons (Clabaugh). The development will consist of 112 one-bedroom units, 306 two-bedroom units, and 98 three-bedroom units ("What is Affordable Housing").

This project is a private-public partnership: funding for this project consists of \$56 million from Amazon's \$2 billion Housing Equity Fund, \$89 million from the Virginia Department of Housing and Community Development, \$227 million from Bank of America, and contributions from the Virginia Housing Trust Fund.

The Exchange will feature a 20,000 square foot community center with an indoor basketball court and fitness center operated by Fairfax County Neighborhood and Community Services. With grocery stores and retail outlets 10 minutes away by foot, as well as public transit options and job opportunities, this development addresses a critical need in a high-opportunity area. Senthil Sankaran, managing principal at the Amazon Equity Housing Fund, emphasized, "As the first 100% affordable multifamily development in Tysons, The Exchange will provide residents with easy access to transit, access to some of the best schools, jobs, and amenities in the region, all in the highly desirable Tysons community." These residential units are even guaranteed to remain affordable for the next 99 years, according to leaders of the Arlington Partnership for Affordable Housing ("What is Affordable Housing").



Fig 1: Image of The Exchange ("What is Affordable Housing")

The Exchange is a prime example of how affordable housing can be successfully integrated into “live, work, play” areas that offer high opportunities. The public-private partnership that helped fund this project highlights the importance of collaborative efforts in tackling the housing crisis. Amazon’s crucial support demonstrates that major corporations can play a role in addressing the needs of communities. Amazon’s Housing Equity Fund alone has boosted Arlington, Virginia’s affordable housing stock by 23% since 2021, proving the impact such corporations can have (“Housing Equity”).

Mosaic District

The Mosaic District in Merrifield, Virginia, is one of Fairfax County’s most popular and social live, work, play communities, providing a mix of residential, commercial, and recreational spaces. Though the exact number of units are not disclosed, there are only a few affordable housing units within the big three luxury buildings: Avalon Mosaic, Modera Mosaic, and Vantage Mosaic. The Mosaic District is conveniently located near bus stops at Lee Highway and Juniper Street (3-minute walk) and Gallows Road (5-minute walk), allowing residents to access job opportunities.

Now, plans for a new development have arisen. Adjacent to the Mosaic District, a vacant Inova office building is being converted into an affordable housing development. Beginning in 2024 and expected to finish in 2025, Fairfax County has approved \$8 million towards this project (Leayman), highlighting the importance of the local government in ensuring high-opportunity areas remain accessible to all residents. Lenore Stanton, chairman of the Fairfax County Redevelopment and Housing Authority, remarked, “Purchasing unused office space in the Mosaic area and converting it into affordable housing will enable individuals and families with lower-income levels to access this highly desirable community.” This new development helps Fairfax County work towards their goal of adding 10,000 additional affordable housing units by 2034. The apartments will be targeted towards households making between 30 and 80 percent of the Area Median Income. For a one-person household, the 80 percent Area Median Income limit is \$66,750. For a four-person household, the limit is \$95,300. Below are the income categories for low, very low, and extremely low-income limits for the DC-VA-MD area in 2023 (HUD).

FY 2023 Income Limit Area	Median Family Income Click for More Detail	FY 2023 Income Limit Category	Persons in Family							
			1	2	3	4	5	6	7	8
Washington-Arlington-Alexandria, DC-VA-MD HUD Metro FMR Area	\$152,100	Very Low (50%) Income Limits (\$) Click for More Detail	52,750	60,300	67,850	75,350	81,400	87,450	93,450	99,500
		Extremely Low Income Limits (\$)* Click for More Detail	31,650	36,200	40,700	45,200	48,850	52,450	56,050	59,700
		Low (80%) Income Limits (\$) Click for More Detail	66,750	76,250	85,800	95,300	102,950	110,550	118,200	125,800

Fig 2: FY 2023 Income Limits Summary (HUD)

Additionally, Merrifield Housing is seeking Virginia Housing for Housing Opportunity Tax Credits for this project. These tax credits incentivize developers to create affordable housing developments by reducing their federal tax liabilities.

This new apartment complex is in a crucial area close to numerous amenities, including supermarkets, quick food options, and schools, ensuring residents have easy access to essentials. In the US, food deserts are a serious issue. According to a report prepared for Congress by the Economic Research Service of the US Department of Agriculture, about 2.3 million people live more than one mile away from a supermarket and do not own a car. This means they have limited access to fresh, nutritious foods, which can lead to poor diets and associated health problems like obesity and diabetes (United States Department of Agriculture). Thus, as the map shows below, the development's (shown as Inova Health System) proximity to amenities such as ALDI and Target help residents prevent the negative effects of food deserts and reduces the financial burden of traveling long distances for groceries.



Fig 3: Map of Mosaic District Development

Location

Although those working in “live, work, play” communities most often cannot afford to live in the very area where they are employed, these communities are essential to the success of affordable housing developments due to their prime location. Both the new affordable housing developments in Tysons Corner and Mosaic District provide numerous benefits. With excellent access to transportation and a wide array of amenities, “live, work, play” communities are location-efficient, making them attractive for affordable housing developments. According to the

U.S. Environmental Protection Agency, housing that are location-efficient are located near public transit, employment centers, and schools.

A key aspect of location efficiency is transportation affordability, which refers to households' ability to move around within their budgets. Financial experts say households should spend no more than 10% of their budgets on transportation (Litman 2). However, households in non-location efficient areas often spend more than 20% of income on transportation. In the past couple years, progress has been made towards faster but more expensive modes, reducing the options for lower-income residents (Hamidi et al.). Addressing transportation affordability in the context of affordable housing is crucial as it directly impacts the cost of living for residents. Researchers have asserted that housing is not truly affordable unless it is located within business hubs and has easy access to affordable transit.

Opposition

As more and more affordable housing developments are initiated across the country, community opposition has become a significant obstacle. Current residents have been expressing concerns that these new developments will alter community culture, increase traffic, and decrease the values of their homes (Ross). A common term people use —“NIMBY” (Not In My Backyard) — reflects their strong resistance to having affordable housing built in their neighborhoods (Pettypiece). However, these concerns are not substantial enough to stop developing affordable housing. Studies have shown little to no evidence to support the claim that affordable housing decreases property values (Nguyen 1). In fact, specific studies were conducted in Alexandria, VA, adjacent to Fairfax County, revealing properties within 1/16 of a mile of a development will **increase** value by 0.09 percent (Stacy et al.). Additionally, the concern about potential change in community culture and character could be viewed positively. As communities become more diverse, they transform into a melting pot of different cultures, perspectives, and experiences.

Solution

In 2020, Fairfax County created an Affordable Housing Task Force to help with recommendations in preserving existing units. To ensure developments remain affordable for many years, they recommended a requirement that requires units to remain affordable for a set number of years, as seen with The Exchange. While the Affordable Housing Task Force has offered strategies such as setting aside a percentage of non-recurring annual balances for affordable housing preservation and offering fee waivers for housing projects, Fairfax County should also consider the following options when expanding affordable housing in the future (Affordable Housing Preservation Task Force).

First, to encourage developers to build more affordable housing units, Fairfax County should expand their use of incentives. Currently, the nationwide Low-Income Housing Tax Credit program reduces federal tax liabilities for developers, and Fairfax County's Tax-Exempt Multifamily Housing Financing Program allows developers to obtain below-market rate

mortgages. However, incentives such as increased floor area ratios and expedited permitting processes can make affordable housing projects more appealing for developers. An increased floor area ratio would mean developers can build more units than permitted by zoning laws if a percentage of the units are designated affordable. Fairfax County has a current maximum floor area ratio of 3.0, meaning the total floor area in a particular building can be up to three times the size of the land itself (“Planned Development District”). Increasing this number will allow developers to increase their profits while still contributing to the community’s affordable housing stock. An expedited permitting process would speed up the approval process for affordable housing projects so that developers can save both time and money. Additionally, public-private partnerships, as seen with The Exchange in Tysons Corner, can bring together resources from different stakeholders, such as the government, non-profits, and corporations like Amazon, to fund and support new developments.

In future developments, location and transportation efficiency must be central considerations. Expanding public transit routes and improving connectivity between affordable housing and major employment hubs, like those in Tysons Corner and Mosaic District, will be essential. While these communities have already integrated affordable housing with access to amenities, further efforts are needed to ensure residents in other areas of Fairfax County are not burdened with high transportation costs, which can undermine the affordability of their housing. One effective approach could be providing subsidies for low-income residents to offset transportation costs. Metro cards, bus passes, and other transit vouchers could be funded by housing authorities and local governments to reduce the financial strain on residents.

Finally, areas in Fairfax County like Reston Town Center and Springfield Town Center should be targeted for future affordable housing developments. These areas have strong economic bases, with numerous job opportunities and access to public transportation, making them ideal for affordable housing. By implementing strategies from Mosaic District and Tysons Corner, Fairfax County can continue making significant strides in addressing its affordable housing needs.

Conclusion

Expanding affordable housing in Fairfax County requires extensive approaches. Strategies such as incentives for developers, public-private partnerships, and offering transportation subsidies are crucial ways for Fairfax County to increase their affordable housing stock, allowing residents regardless of economic status to integrate into high-opportunity areas. As one of the largest and established counties in the country, Fairfax County has a responsibility to lead by example. By implementing these strategies and committing to long-term affordability, Fairfax County, and soon other counties across the nation, can continue to progress while ensuring all community members have access to homes.

Works Cited

- Affordable Housing Preservation Task Force. *Recommendations to the Fairfax County Board of Supervisors*. 6 Apr. 2021.
- Clabaugh, Jeff. "Fairfax County Breaks Ground on Tysons' First 100% Affordable Residential Development." *WTOP News*, 26 Dec. 2023, [wtop.com/business-finance/2023/12/fairfax-county-breaks-ground-on-tysons-first-100-affordable-residential-development/](https://www.wtop.com/business-finance/2023/12/fairfax-county-breaks-ground-on-tysons-first-100-affordable-residential-development/).
- "§ 15.2-2201. Definitions." *LIS Virginia Law*, law.lis.virginia.gov/vacode/title15.2/chapter22/section15.2-2201/. Accessed 18 Sept. 2024.
- "Fairfax, VA Rental Market Trends." *Average Rent in Fairfax, VA & Rent Price Trends*, Apartment List, www.apartmentlist.com/rent-report/va/fairfax. Accessed 16 Sept. 2024.
- Hamidi, Shima, et al. "Does location matter? performance analysis of the affordable housing programs with respect to transportation affordability in Dallas Fort Worth (DFW) metropolis." *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 3, 8 Aug. 2018, pp. 194–205, <https://doi.org/10.1177/0361198118790838>.
- "Housing Equity." *US About Amazon*, Amazon, 6 Jan. 2021, www.aboutamazon.com/impact/community/housing-equity.
- Kimura, Donna. "Pop Quiz with Carmen Romero." *Housing Finance*, 21 Mar. 2024, www.housingfinance.com/news/pop-quiz-with-carmen-romero_o.
- Leayman, Emily. "Affordable Housing Converted from Office Planned near Mosaic District." *Falls Church, VA Patch*, Patch, 3 Aug. 2023, patch.com/virginia/fallschurch/affordable-housing-converted-office-planned-near-mosaic-district.
- Leayman, Emily. "Affordable Housing Converted from Office Planned near Mosaic District." *Falls Church, VA Patch*, Patch, 3 Aug. 2023, patch.com/virginia/fallschurch/affordable-housing-converted-office-planned-near-mosaic-district.
- Litman, Todd. "Transportation Affordability Evaluation and Improvement Strategies." *Victoria Transport Policy Institute*, 17 Dec. 2021, www.vtpi.org/affordability.pdf.
- Mary Kate McGrath. "Everything Developers Need to Know about Live Work Play Communities." *Northspyre.com*, Northspyre, 9 Apr. 2024, www.northspyre.com/blog/everything-developers-need-to-know-about-live-work-play-communities. Accessed 23 Sept. 2024.
- Nguyen, Mai Thi. "Does Affordable Housing Detrimentially Affect Property Values? A Review of the Literature." *Journal of Planning Literature*, vol. 20, no. 1, Aug. 2005, pp. 15–26, <https://doi.org/10.1177/0885412205277069>.
- Pettypiece, Shannon. "In 'warfare against Renters,' Homeowners Fight Affordable Housing Push." *NBC News*, NBC Universal News Group, 17 July 2023,

- www.nbcnews.com/politics/economics/housing-prices-surge-rents-homeowners-fight-rent-a93789.
- “Planned Development District Regulations.” *Fairfax County*,
www.fairfaxcounty.gov/tysons/sites/tysons/files/assets/documents/pdf/zoning/art06.pdf.
Accessed 16 Sept. 2024.
- Ross, Jaimie, and Kody Glazer. *Overcoming NIMBY Opposition to Affordable Housing*.
National Low Income Housing Coalition, 2022.
- Stacy, Christina Plerhoples, and Christopher Davis. “Assessing the Impact of Affordable
Housing on Nearby Property Values in Alexandria, Virginia.” *Urban Institute*, 19 Apr.
2022,
www.urban.org/research/publication/assessing-impact-affordable-housing-nearby-property-values-alexandria-virginia#
- United States Department of Agriculture Economic Research Service, 2009, *Access to Affordable
and Nutritious Food: Measuring and Understanding Food Deserts and Their
Consequences.*,
https://www.ers.usda.gov/webdocs/publications/42711/12716_ap036_1_.pdf?v=41055.
- “What Is Affordable Housing?” *APA*, 19 July 2022,
www.apah.org/housing/what-is-affordable-housing/.