

|CURIEUX|

ACADEMIC JOURNAL

November Issue

Part 2 Issue 1

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 Association Between School-Related Perceived Financial Pressure and Adolescent Mental Well-Being By Cynthia Zhang

Page 19: Redefining Richard Florida's Creative Class (CC): Artificial Intelligence and the Future of the Creative & Cultural Industries (CCIs) By Peri Yeung

Page 39: Designing Inclusive Digital Platforms for the Silver Generation By Sehyeon Kim

Page 54: Evaluating Psychedelics in Depression Treatment: A Comparative Review of Ketamine, Mescaline, and Psilocybin By Hanayo Endo

Page 68: Is Objectivity All in the Mind? Investigating the Psychological Construction of Health By Kiyan Kapur

Page 82: Factors that Influence Youth Perceptions of the Electoral College By Tomer Fine

Page 103: Calabasas Students Perceptions of Pretty Privilege By Ella Meyers

Page 121: PhantomRx: Alleviating Phantom Limb Pain Using Bionic Hand By Ansh Saraf, Ti Air Riggins

Page 131: The Impact of Gloomy and Sunny Conditions on AI-Based Rip Current Detection By Chilok Mak

Page 141: Hierarchical Label Structure-based Curriculum Learning for Lung and Colon Histopathological Image Classification By Junahn Choi

Page 152: Domination and Disavowal: Israel, Palestine, and the Question of the Settler-Colonial Framework By Ariya Harrington

Page 161: A Comprehensive Review of Artificial Intelligence Applications in the Stock Market By Lindsey Mills

Page 170: FinCoach: A Lightweight AI-Powered Financial Coaching System for Students By Minh Duc Dang Vu

Page 182: Crisis Response, Pricing Decisions and Global Trends: A Micro and Macro-Economic Study of the Food and Beverage Industry By Vivan Agarwal

Page 193: A Review of Recurrent Architectures for Human Action Recognition By Arjav Jain

Page 209: How Mesenchymal Stem Cells Play a Role in Different Injuries By Viktoria Lastivka

Page 223: "Very nearly... Equal in Status to Caucasians:" The Mississippi Delta Chinese's Social and Racial Transformation (1890-1960) By Brendan Wong

Page 234: Gender Inclusion in STEM: Feminist Perspectives on VEX Robotics By Anya Li

Page 246: A Study of AI's Impact on Education: Evaluating the Economic Sustainability and Demand for Traditional Tutoring Services In the US By Aaryan Kumar

Page 265: Breaking the Biofilm Barrier: Tackling Antibiotic-Resistant Infections with Gene Editing, Quorum Sensing, and AI-Driven Solutions By Finneas Murphy

Page 275: Brain Regions and Neural Networks that Encode Semantic Memory By Milton Dai

Page 292: Analyzing the Russia-Ukraine Crisis's Impact on the Global Trade of Agricultural Products and Energy Resources By Syaon Kumar

Page 300: Exploring the Relationship Between United States Electric Vehicle Demand and Taiwanese Manufacture of Integrated Circuits By Joshua Liu

Page 307: Sociodemographic Predictors of Delayed or Unmet Medical Care Among California Adolescents By Lauren Chen

Page 316: RAPIDSS: Rat brain Automated Pipeline for DCE-MRI Segmentation and Signal-analysis By Liana Kim

Page 327: PARO Robot in Dementia Care: A Systematic Review With Narrative Synthesis By Hillary Liu Sun

Page 340: Feminism: Styled, Scripted, and Syndicated By Tanisha Chhabra

Page 346: Correlation Between the Common Psychological Responses Athletes Experience Post-injury By Yeonjae Kim

Page 354: The Effects of Message Framing on Charitable Donations From High School Students By Ketan Shah

Page 361: CRISPR Cas-9 as a Potential Cure for Wilson's Disease By Satoshi Toya

Page 397: Legal Battles for Belonging: Chinese Immigrants and the Constitution, 1886–1927 By Tianli Cheng

Page 404: Navigating Economic Recovery Post-COVID-19: The Role of Policy, Labor Flexibility, Healthcare, and Digital Infrastructure By Shreemay Shah

Page 417: Evaluating the Optimal Machine Learning Model for Renewable Energy Source Selection Based on Environmental and Regional Attributes By Kelvin Zebah

Page 425: A Comparative Analysis of User Growth: Understanding TikTok's Follower Dominance Over Little Red Note By Cindy Wu

Page 431: Unpacking the Economic and Sociocultural Roots of South Korea's Fertility Collapse By Claire Yura Kim

Page 458: Diving into Innovation: Utilizing Kingfisher and Gannet Adaptations for Rocket Nose Cone Design By Lauren Johnson

Page 478: "His Family Tree Recedes": A Personal Narrative Analysis Discussing Dementia Caregiving Themes, Strategies, and AI Implementation By Noel Yi

Page 500: Renal Cell Carcinoma: Epidemiology, Risk Factors, Genetics, and Treatment. By Sabrina Kainbacher

Page 508: How BRICS CBDC can Reduce Dollar Dominance By Tariq Murgian

Page 523: Exploring the Effects of Stress, Anxiety, and Depression on Cognition, and the Therapeutic Influence of Music By Tamira Darisala By Mentored and by Dr. Deblina Dey

Page 533: Ultraviolet Vision in *Heliconius* Butterflies: From Opsin Duplication to Behavioral Ecology By Max Gan

Page 538: Whether Corporate Governance in India Post Liberalization and Colonialism Shows No Distinction from Pre Liberalization and Colonialism By Trayambika Reddy Chirra

Page 552: AI-Powered Early Warning System: Safeguarding Microenterprises Through Predictive Analytics By Vrinda Bisoi

Page 567: Rate, Load, and Medium Effects on Hysteresis in Coiled NiTi Springs By Agastyaa Poddar

The Association Between School-Related Perceived Financial Pressure and Adolescent Mental Well-Being By Cynthia Zhang

Abstract

This study investigated the relationships between adolescents' school-related perceived financial pressure, mental well-being, and their relationship with their parents. A 27-item online questionnaire was designed with three subscales - perceived financial pressure, parent-adolescent relationship, and mental well-being - to examine this. A sample of 40 adolescent students in Auckland, New Zealand, was collected. Correlational relationships between the subscales were measured using Pearson's correlation, and gender differences were investigated using the ANOVA test. The correlational tests revealed that there are significant negative correlations between perceived financial pressure and parent-adolescent relationship, as well as between perceived financial pressure and mental well-being. A positive correlation was identified between parent-adolescent relationship and mental well-being. No significant differences within these subscales were found between genders. The findings suggest that adolescent students' stress related to academics is significantly influenced by how they perceive their parents' financial situation. It is also suggested that financial stress places a strain on parent-adolescent relationships, as consistent with previous research. The results also give reason to assume that male and female adolescents in Auckland experience similar amounts of financial pressure.

Introduction

Adolescence is a critical period of identity development, which is greatly influenced by the social circle surrounding them. It is a time spent largely on experiences related to school education. As a result, academic pressure, which can be defined as stress that stems from one's academic performance (Lan et al.), is found to be one of the leading factors that determine the state of an adolescent's mental well-being (Steare et al.). Specifically, numerous studies have identified academic pressures as the most influential causes of perceived stress (Dopmeijer; Neveu et al.; Mitchell). This is a serious issue as adolescence is represented by an increased sensitivity to stress. Consequently, adolescents are susceptible to mental health issues (Baik et al.; Alharbi and Smith). Research has discovered that increased academic pressure is associated with increased depressive and anxiety symptoms (Tang et al.; Steare et al.). Additionally, research discovered a strong positive correlation between perceived stress and the female gender (Farrer et al.), hence suggesting that females may be more mentally affected by academic pressures than males. The delicate nature of mental well-being for individuals at this age results in a need to dive deeper into the nuance of the broad term, academic pressure. This research will help discover solutions that alleviate the intensity of academic pressure and hence its negative effects on mental health. Central factors that will be investigated in this study are parental pressure and financial pressure.

The perceived parental support or pressure in regards to academics plays a key role in an adolescent's development of stress (Kulakow et al.). Parental support can be viewed as a

resource, hence, according to Hobfoll's (1989) definition of psychological stress as "a reaction to the environment in which there is (a) a threat to a net of resources, (b) the net loss of resources, or (c) a lack of resource gain following the investment of resources", the opposite of parental support - parental pressure - can be deemed a source of stress. Parental pressure being a stressor is supported by past research that has found that it positively correlates with anxiety (Ringeisen and Raufelder). This includes general anxiety and test anxiety, which is when an individual experiences extreme stress and anxiety before and during a test (Stankovska et al.). Intriguingly, while girls will view parental support and pressure as opposites, boys are more likely to see them as the same thing (Leff and Hoyle; Raufelder et al.). This can be explained by the idea that parental pressure and parental support, specifically informational support and not emotional, are opposite ends of the same spectrum - they both involve conveying interest in the child's academic outcomes (Ringeisen and Raufelder).

In addition to parental pressure, the perceived stress of college students has a strong correlation with their financial situation (Britt et al.). Financial troubles may be a direct contributor to a self-providing individual's academic stress, but for adolescents, it is more likely that it indirectly affects academic pressure by influencing the levels of perceived parental pressure. Economic stress increases parents' emotional distress, decreasing parental warmth and hence decreasing the quality of parent-child relationships (Xiao et al.). However, there is a scarcity of research that directly investigates the effect of parents' financial stress on adolescents' sense of academic pressure.

This study aims to investigate the extent to which parents' financial stress and investment in their New Zealand adolescent child influence the parent-adolescent relationship and adolescents' mental well-being in regard to academic pressure. Given the sensitive nature of adolescents' mental health and the significance of school-related experiences to teenagers, this research is timely. Our hypotheses include: (1) levels of perceived financial pressure will negatively correlate with adolescent school-related mental health, (2) the quality of the parent-adolescent relationship will positively correlate with school-related mental well-being, and (3) levels of perceived financial pressure will negatively correlate with the quality of the parent-adolescent relationship. This study will also test the hypothesis that females will generally have worse mental well-being and increased perceived financial pressure.

Materials and Methods

Participants and procedure

This study used a convenience sampling method to collect data from 40 adolescent students in Auckland, New Zealand. The participants were 15-19 years old (mean = 17.25, SD = 0.776). 13 participants were male (32.5%), and the remaining 27 participants were female (67.5%). 60% of the participants were Chinese, 20% Korean, and 13% White/Caucasian. One participant was multiracial, and two were of other races. Most of the participants reported their family's socioeconomic status as middle-class (n = 28); six were of high socioeconomic status,

and 1 was low-class. The majority of the participants were second-generation immigrants to New Zealand (60%). 33% were first-generation immigrants, and the rest were 3rd or more generations immigrants.

This study followed the ethical standards of a private school system in Auckland, New Zealand. All participants were asked to complete a consent form, which also contained information regarding the details and purpose of the study. The privacy of the participants was ensured by designing the survey questionnaire responses to be anonymous; consent was signed on a separate form so that the names of the participants could not be connected to their responses. No data that could be used to identify individuals was collected. Additionally, all participants were reminded of their right to withdraw before starting the survey.

Measures

A 27-item online questionnaire (Appendix A) was designed specifically to test the hypotheses of this study. It included a background section, along with 3 separately scored sections and an optional open question. Each item in the scored sections was individually scored on a 4-point Likert scale (1-4). The background section involved 5 items that collected information on the participants' demographic: age, gender, ethnicity, socioeconomic status and immigration status. The scored sections were composed of the 3 following domains: perceived financial pressure, parent-adolescent relationship and mental well-being. The final open question was: "How do your parents' expectations affect your mental well-being, if at all?"

Perceived financial pressure. This subscale of the questionnaire was made up of 6 items that measured the adolescent's perception of their parents' general financial stress. It also measured the adolescent's perception of their parents' financial investment in school-related activities, expectations for the child's future repayment, and influence on the adolescent's decisions. An example of a question is: "To what extent does your parents' financial investment in an activity influence your decisions related to said activity?" The minimum score that could be achieved for this section was 6, and the maximum was 24. A higher score represented a higher level of perceived financial pressure. This section also included a follow-up item that the participant would be required to answer depending on their response to the previous question, "How often do you consider your parents' financial contributions when making decisions?" This follow-up question was not included in the scoring, as it collected qualitative answers that did not lie on a scale. The follow-up item was: "Which word best describes why you consider your parents' financial contributions?", with the answer options of: "Gratitude", "Obligation", "Dependence", and "Guilt".

Parent-adolescent relationship. This subscale consisted of 7 scored items and 1 non-scored item ("I always do what my parents tell me to"). The minimum score that could be obtained for this subscale was 7, and the maximum was 28. A higher score would correspond with worse parent-adolescent relationship quality. In this study, a lowered parent-adolescent relationship quality is characterised by increased general parent strictness, influence and pressure on the child, along with a decreased sense of openness from the child to the parent.

Mental well-being. Mental well-being was measured by 7 items, again all scored on a 4-point Likert scale. This subscale focused on general stress, stress related to school, uncertainty and overall life satisfaction. A question from this section is “How stressed do you get during exam periods due to pressure from parents to do well?” The participants could score a minimum of 7 and a maximum of 28. The higher the score, the worse the participant’s mental well-being.

Data analysis

The digital analysis program, JASP, was used to analyse the data collected by the questionnaire. Scores collected from each subscale of the questionnaire (perceived financial pressure, parent-adolescent relationship, mental well-being) were separated into three separate data sets. These three sets of data were first tested to determine which correlation test would be suitable. They were plotted into scatter graphs to confirm that they had linear relationships with one another. Additionally, a Shapiro-Wilk value was calculated for each subscale to test whether the data within each set was normally distributed. This test was used as it is suitable for data sets of samples smaller than 50. Concluded from the information gathered in the above analysis, Pearson’s correlation was used to analyse the correlational relationships between the subscales. The following correlational relationships were investigated to test the hypotheses: perceived financial pressure v parent-adolescent relationship, perceived financial pressure v mental well-being, and parent-adolescent relationship v mental well-being. The p-values obtained from each Pearson’s r-value will be compared with the alpha level of 0.05; in other words, a p-value smaller than 0.05 will indicate a significant correlation.

In addition to correlational tests between subscales, the differences in data between males and females will be investigated. A one-way ANOVA test method will be used to determine whether there is a significant difference in the data collected within each subscale between the different genders. The Brown-Forsythe test was first used to confirm the homogeneity of variance between data sets. Like with the correlational tests, an alpha level of 0.05 will be established for the ANOVA tests.

Results

Descriptive Statistics

Table 1. Descriptive Statistics

		Parent-Adolescent Relationship	Mental Well-being
Valid	40	40	40
Mean	13.275	15	17.425
Std. Deviation	3.146	3.187	3.186
Shapiro-Wilk	0.962	0.947	0.962

P-value of Shapiro-Wilk	0.191	0.06	0.192
Minimum	8	7	12
Maximum	21	23	25

Table 1 presents the descriptive statistics of each data set of the three subscales within the questionnaire. This includes the mean, standard deviation, and Shapiro-Wilk test values. The p-values of the Shapiro-Wilk test for each data set were 0.191, 0.060, and 0.192 for the variables perceived financial pressure, parent-adolescent relationship, and mental well-being, respectively. Since all the p-values were above 0.05, the results for the Shapiro-Wilk test were not significant, and hence the data within each subscale can be assumed to be normally distributed. This confirms that Pearson's correlation can be used to analyse the relationship between the subscales.

Correlations

Table 2. Pearson's Correlations

			Pearson's r	p	Effect size (Fisher's z)	SE Effect size
Perceived Financial Pressure	-	Parent-Adolescent Relationship	0.35*	0.027	0.366	0.164
Perceived Financial Pressure	-	Mental Well-being	0.687***	< .001	0.841	0.164
Parent-Adolescent Relationship	-	Mental Well-being	0.437**	0.005	0.469	0.164
* p < .05, ** p < .01, *** p < .001						

Table 2 displays the results of Pearson's Correlational tests. This includes Pearson's r value, the p-value, and the effect size (Fisher's z) of the three pairs of data sets. All relationships had a p-value smaller than 0.05, so they were all significant. Additionally, a positive correlation was found for each relationship. The relationship between perceived financial pressure and mental well-being was the most positively associated ($r = 0.687$, $p < 0.001$), and had the largest effect size ($z = 0.841$). Parent-adolescent relationship and mental well-being had a smaller positive association ($r = 0.437$, $p = 0.005$) with a moderate effect size ($z = 0.469$). Perceived financial pressure and the parent-adolescent relationship were most weakly positively related ($r = 0.350$, $p = 0.027$) and had the smallest effect size ($z = 0.366$).

Gender

Table 3. ANOVA - Perceived Financial Pressure

Cases	Sum of Squares	df	Mean Square	F	p

Gender	3.354	1	3.354	0.333	0.567
Residuals	382.621	38	10.069		

Table 4. ANOVA - Parent-Adolescent Relationship

Cases	Sum of Squares	df	Mean Square	F	p
Gender	2.849	1	2.849	0.275	0.603
Residuals	393.151	38	10.346		

Table 5. ANOVA - Mental Well-Being

Cases	Sum of Squares	df	Mean Square	F	p
Gender	2.333	1	2.333	0.225	0.638
Residuals	393.442	38	10.354		

Within each subscale, the data were separated into two groups: male and female. The results of the Brown-Forsythe test for equality of variance between genders in each subscale all produced a p-value larger than 0.05 (perceived financial pressure, $p = 0.644$; parent-adolescent relationship, $p = 0.611$; mental well-being, $p = 0.889$); hence, the variances within each subscale can be assumed to express homogeneity. Table 3, table 4, and table 5 present the ANOVA test results for the perceived financial pressure, parent-adolescent relationship, and mental well-being subscales, respectively. There is no statistically significant difference between genders in regards to their scores in the perceived financial pressure subscale ($F = 0.333$, $p = 0.567$), as well as for the parent-adolescent relationship ($F = 0.275$, $p = 0.603$) and mental well-being ($F = 0.225$, $p = 0.638$) subscales.

Qualitative data

The non-scored follow-up item under the financial pressure subscale had the options of “guilt”, “gratitude”, “obligation”, and “dependence”. Only two participants were not given the follow-up question, as they answered “Never” to the previous question asking “How often do you consider your parents’ financial contributions when making decisions?” Out of the 38 participants that answered the follow-up, 25 answered “gratitude”, 3 answered “obligation”, 3 answered “dependence”, and 5 answered “guilt”.

25 participants answered the optional open question, “How do your parents’ expectations affect your well-being, if at all?” 14 (56%) of these answers were interpreted as indicating that the parents had no significant effect on the adolescent’s well-being. The following are examples of some of the answers that were interpreted this way: “it doesn’t”, “a little bit”, “They [...]

encourage mental wellbeing and overall happiness [over] the pressures and stress that comes from excessive study and academic work”, “They want for me what I want for myself. I don’t feel that they have unrealistic standards.” 6 (24%) participants expressed that their parents’ expectations had a significant effect on their mental well-being, for example. Of these participants, one said that these impactful expectations were on their sexuality and marriage rather than academics. 5 (20%) participants felt that their own expectations had a greater impact on their mental well-being than their parents’ expectations. Out of all 25 participants, 4 (16%) expressed that their mental well-being is impacted by their own desire to repay their parents: “Sources of stress due to their expectations would rather be due to me feeling the need to ‘repay’ their efforts in funding my education”, “If I don’t get a good job in the future, all the money and time they invested into me is going to be useless”, “my mum’s a really hard worker and I would just like to live up to her expectations as she put her all into raising me”, “[...] I think I’m harder on myself, knowing they have spent so much money and time on me.” Interestingly, of these 4 participants, 3 answered “guilt” in the follow-up question while the other responded with “gratitude”.

Discussion

This study aimed to investigate the extent to which parents’ financial stress and investment in their child, as perceived by the adolescent child, influences the parent-adolescent relationship and the adolescent’s mental well-being. This study’s 3 main hypotheses were all confirmed by the results. Hypothesis 1 assumes that levels of perceived financial pressure will negatively correlate with adolescent mental health, meaning that as perceived financial pressure increases, adolescent mental health will worsen. This hypothesis was confirmed as the scores on the perceived financial pressure and mental health subscales had a positive correlation. Since a higher score on each subscale equated to increased pressure and worse mental health, respectively, the positive correlation of the subscale scores corresponds to a negative correlation between perceived financial pressure and adolescent mental health, thus confirming hypothesis 1. This is consistent with previous research that found financial pressure in childhood can affect the individual on a cellular level by causing shortened telomeres, which is associated with psychological illnesses and traumas (Tung et al.). Numerous studies also established a positive correlation between familial financial pressure and a child’s socioemotional difficulties, depression and suicidal ideation (McGill et al.; Yang et al.).

Hypothesis 2 assumes that the quality of the parent-adolescent relationship will positively correlate with mental well-being. This hypothesis was confirmed by the positive correlation that was found between the parent-adolescent and mental well-being subscales. This is, again, consistent with past research that established sub-optimal parenting would result in the child having worse mental health (Stafford et al.). An explanation for this correlation could be that children have an increased sense of loneliness when their relationship with their parents lacks closeness (Yan et al.). This loneliness implies an absence of a sense of safety, which is a consequence of the inadequate emotional and physical care received by the child from the parent.

The child is deprived of this opportunity to form emotional connections, which are necessary for developing healthy coping strategies for stressful situations (Yang et al.).

Finally, our 3rd main hypothesis - levels of perceived financial pressure will negatively correlate with the quality of the parent-adolescent relationship - was also proven correct. This hypothesis assumes that an increase in financial pressure will be accompanied by a decrease in the quality of the parent-adolescent relationship. The results found a positive correlation between financial pressure and parent-adolescent relationship subscales' scores, where a higher score in each meant increased financial pressure and decreased quality of parent-adolescent relationship, respectively. Hence, the positive correlation established by the scores corresponded to the relationship assumed in the hypothesis. The perceived financial pressure reported by adolescents is dependent on how their parents present their financial stress. Hence, adolescents would likely experience an increased perceived financial pressure if their parents themselves were experiencing more financial stress. Economic hardships make parents vulnerable to psychological distress, as proven by their positive correlation with depression (Guan et al.). Parents experiencing more mental distress is associated with increased neglect of their child and maladaptive parenting styles (Dadi et al.; Tung et al.).

Although all these relationships were established separately, they can be considered in tandem with each other. Parental financial worry can increase psychological stress which may be reflected by negative parent-child interactions, hence decreasing the quality of the parent-adolescent relationship. These negative interactions, such as neglect and oppressive control, will result in the child feeling an absence of security and, consequently, developing higher levels of anxiety (Thomson et al.). This explanation is supported by studies that found associations between family economic stress, parent-child conflicts, and decreased mental well-being of the child (e.g., Mistry and Elenbaas).

The hypothesis regarding gender, which assumed females would have worse mental well-being and increased financial pressure, was disproven by the results. No significant statistical differences were found in the scores for the mental well-being, financial pressure, and parent-adolescent relationship subscales between genders. As a result, we can conclude that male and female adolescents are equally susceptible to financial pressures and mental distress. This is inconsistent with previous research that has established a general consensus that women are more vulnerable to worsened mental health, such as increased levels of stress, due to higher sensitivity to external emotional stressors (e.g., Tung et al.; Barbayannis et al.). However, other studies had findings that could be interpreted as contradicting this. For example, academic burnout prevalence has no significant associations with gender (Shrestha et al.). This study's results can be explained by considering how different genders will interpret external stimuli in different ways. Friends and peers begin to have a larger influence on an individual's mental well-being, while the family's role diminishes. This effect occurs earlier in girls (age 12) than in boys (age 16) (Cilar Budler and Stiglic). The effect of adolescent girls' general increased sensitivity to familial stress can be assumed to be dampened by their earlier decreased need for

parental support compared to boys. In conclusion, the different factors influencing mental well-being balance each other out, resulting in no significant difference between genders.

The qualitative data collected in this study give insight into the well-being and thoughts of the adolescent participants. From this data, it can be assumed that over 50% of the participants felt that their parents' expectations did not have a large impact on their mental well-being. The rest of the participants were divided roughly in half, where one group felt that their parents' expectations had a significant effect, while the other group felt that the effects of their own expectations outweighed those of their parents. This shows that there are more adolescents in Auckland, New Zealand, who don't feel overwhelmed by parental expectations than adolescents who do. A small percentage of participants (around 16%) felt that they needed to prove that the money invested in them by their parents was worth it. It can be assumed that the majority of these participants felt this way due to guilt for not living up to the standards set by the money spent. Since these participants specified that their stress was rooted in the financial investment rather than the parents' explicit expectations, a conclusion can be drawn that financial investment is of the same domain as financial pressure. Thus, parents' financial investment indirectly contributes to academic pressures in adolescents.

Limitations

Some limitations should be considered when taking in the findings of this study. Firstly, the questionnaire collects data in the form of self-reported information. As a result, the participants' responses are subjected to recall biases. However, a self-report was deemed an appropriate and beneficial method, as it allowed for the assessment of the adolescents' internal state and perception. Secondly, this study involved a small sample size of only 40 participants, meaning that it is possible some results were produced by chance. Hence, the results of this study may not be entirely representative of the adolescent community in New Zealand. Lastly, a convenient sampling method was implemented. This resulted in a skewed sample in terms of gender, ethnicity, socioeconomic status, and immigration status. The majority of participants were female, Chinese, middle-class, and 2nd generation immigrants.

Future directions

The current findings describe the effects of perceived financial pressure on adolescent well-being and parent-child relationships. These suggest a need for parents to adjust how they express their financial stress at home in order to maintain their adolescent child's mental health. To help parents better understand what specific areas they should consider when interacting with their child, future studies should examine what factors may contribute to an adolescent's perceived financial pressure. For example, future research should attempt to identify the different effects of financial investment, parental financial stress, or wants versus needs (Mistry and Elenbaas). A focus should also be placed on discovering the emotions that underlie an adolescent's financial pressure.

The negative correlation between perceived financial pressure and parent-child relationship quality indicates that financial stress leads to family stress (Friedline et al.; Rodrigues et al.). This strain on familial relationships likely arises from dysfunctional parenting practices (Dadi et al.). Future studies should define distinct aspects of dysfunctional parenting styles and identify which affect child mental well-being most significantly. These aspects may include physical abuse, neglect, parental perfectionism, and harsh parenting (Dadi et al.; Friedline et al.; Streit et al.; Randall et al.).

Future studies may also focus on gender differences regarding financial pressure, parent-child relationships, and well-being related to academics. The ambivalent nature of the results drawn from past research and this current study indicates a need for research to be done in greater detail. Researchers should focus on identifying reasons for gender differences, or lack thereof, that also align reasonably with conclusions from previous research.

Works Cited

- Alharbi, Eman, and Andrew Smith. "Review of the Literature on Stress and Wellbeing of International Students in English-Speaking Countries." *International Education Studies*, vol. 11, no. 6, May 2018, p. p22. www.ccsenet.org, <https://doi.org/10.5539/ies.v11n6p22>.
- Baik, Chi, et al. "How Universities Can Enhance Student Mental Wellbeing: The Student Perspective." *Higher Education Research & Development*, vol. 38, no. 4, June 2019, pp. 674–87. Taylor and Francis+NEJM, <https://doi.org/10.1080/07294360.2019.1576596>.
- Barbayannis, Georgia, et al. "Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19." *Frontiers in Psychology*, vol. 13, 2022, p. 886344. PubMed, <https://doi.org/10.3389/fpsyg.2022.886344>.
- Britt, Sonya L., et al. "Student Loans, Financial Stress, and College Student Retention." *Journal of Student Financial Aid*, vol. 47, no. 1, Apr. 2017. DOI.org (Crossref), <https://doi.org/10.55504/0884-9153.1605>.
- Cilar Budler, Leona, and Gregor Stiglic. "Age, Quality of Life and Mental Well-Being in Adolescent Population: A Network Model Tree Analysis." *Scientific Reports*, vol. 13, Oct. 2023, p. 17667. PubMed Central, <https://doi.org/10.1038/s41598-023-44493-w>.
- Dadi, Dunia, et al. "Emerging Adult College Students' Descriptions of Exposure to Childhood Emotional Abuse and Associated Factors: A Qualitative Exploration." *Child Abuse & Neglect*, vol. 146, Dec. 2023, p. 106516. PubMed, <https://doi.org/10.1016/j.chiabu.2023.106516>.
- Dopmeijer, Jolien. *Running on Empty: The Impact of Challenging Student Life on Wellbeing and Academic Performance*. Jolien Dopmeijer, 2021.
- Farrer, Louise M., et al. "Demographic and Psychosocial Predictors of Major Depression and Generalised Anxiety Disorder in Australian University Students." *BMC Psychiatry*, vol. 16, no. 1, July 2016, p. 241. BioMed Central, <https://doi.org/10.1186/s12888-016-0961-z>.
- Friedline, Terri, et al. "Families' Financial Stress & Well-Being: The Importance of the Economy and Economic Environments." *Journal of Family and Economic Issues*, vol. 42, no. Suppl 1, 2021, pp. 34–51. PubMed Central, <https://doi.org/10.1007/s10834-020-09694-9>.
- Guan, Naijie, et al. "Financial Stress and Depression in Adults: A Systematic Review." *PloS One*, vol. 17, no. 2, 2022, p. e0264041. PubMed, <https://doi.org/10.1371/journal.pone.0264041>.
- Hobfoll, Stevan E. "Conservation of Resources: A New Attempt at Conceptualizing Stress." *American Psychologist [US]*, vol. 44, no. 3, 1989, pp. 513–24. APA PsycNet, <https://doi.org/10.1037/0003-066X.44.3.513>.
- Kulakow, Stefan, et al. "School-related Pressure and Parental Support as Predictors of Change in Student Stress Levels from Early to Middle Adolescence." *Journal of Adolescence*, vol. 87, no. 1, Feb. 2021, pp. 38–51. Crossref, <https://doi.org/10.1016/j.adolescence.2020.12.008>.
- Lan, Zhensong, et al. "The Impact of Academic Pressure and Peer Support on Adolescents' Loneliness: A Polynomial Regression and Response Surface Analysis." *Psychology*

- Research and Behavior Management, vol. 16, Nov. 2023, pp. 4617–27. PubMed Central, <https://doi.org/10.2147/PRBM.S435977>.
- Leff, Stephen S., and Rick H. Hoyle. “Young Athletes’ Perceptions of Parental Support and Pressure.” *Journal of Youth and Adolescence*, vol. 24, no. 2, Apr. 1995, pp. 187–203, <https://doi.org/10.1007/BF01537149>.
- McGill, Megan G., et al. “Financial Stress during COVID-19: Implications for Parenting Behaviour and Child Well-Being.” *BMJ Paediatrics Open*, vol. 6, no. 1, Sept. 2022, p. e001569. PubMed, <https://doi.org/10.1136/bmjpo-2022-001569>.
- Mistry, Rashmita S., and Laura Elenbaas. “It’s All in the Family: Parents’ Economic Worries and Youth’s Perceptions of Financial Stress and Educational Outcomes.” *Journal of Youth and Adolescence*, vol. 50, no. 4, 2021, pp. 724–38. PubMed Central, <https://doi.org/10.1007/s10964-021-01393-4>.
- Mitchell, Andrew E.P. “The Perceived Psychological Stressors and Coping Behaviours in University Students, on a Pre-Registration Programme.” *The Journal of Mental Health Training, Education and Practice*, vol. 15, no. 4, July 2020, pp. 249–59. Silverchair, <https://doi.org/10.1108/JMHTEP-09-2019-0048>.
- Neveu, D., et al. “Students Perceived Stress in Academic Programs: Consequences for Its Management.” *Revue d’Épidémiologie et de Santé Publique*, vol. 60, no. 4, Aug. 2012, pp. 255–64. Semantic Scholar, <https://doi.org/10.1016/j.respe.2012.01.008>.
- Randall, Edin T., et al. “Understanding Affluent Adolescent Adjustment: The Interplay of Parental Perfectionism, Perceived Parental Pressure, and Organized Activity Involvement.” *Journal of Adolescence*, vol. 41, no. 1, June 2015, pp. 56–66. Crossref, <https://doi.org/10.1016/j.adolescence.2015.03.005>.
- Raufelder, Diana, et al. “The Perceived Role of Parental Support and Pressure in the Interplay of Test Anxiety and School Engagement Among Adolescents: Evidence for Gender-Specific Relations.” *Journal of Child and Family Studies*, vol. 24, no. 12, Dec. 2015, pp. 3742–56. Springer Link, <https://doi.org/10.1007/s10826-015-0182-y>.
- Ringeisen, Tobias, and Diana Raufelder. “The Interplay of Parental Support, Parental Pressure and Test Anxiety – Gender Differences in Adolescents.” *Journal of Adolescence*, vol. 45, no. 1, Dec. 2015, pp. 67–79. Crossref, <https://doi.org/10.1016/j.adolescence.2015.08.018>.
- Rodrigues, Margarida, et al. “COVID-19: Financial Stress and Well-Being in Families.” *Journal of Family Issues*, vol. 44, no. 5, May 2023, pp. 1254–75. PubMed Central, <https://doi.org/10.1177/0192513X211057009>.
- Shrestha, Dhan Bahadur, et al. “Burnout among Medical Students of a Medical College in Kathmandu; A Cross-Sectional Study.” *PloS One*, vol. 16, no. 6, 2021, p. e0253808. PubMed, <https://doi.org/10.1371/journal.pone.0253808>.
- Stafford, Mai, et al. “Parent-Child Relationships and Offspring’s Positive Mental Wellbeing from Adolescence to Early Older Age.” *The Journal of Positive Psychology*, vol. 11, no. 3, May 2016, pp. 326–37. PubMed, <https://doi.org/10.1080/17439760.2015.1081971>.
- Stankovska, Gordana, et al. “BCES Conference Book.” *Emotional Intelligence*, vol. 16, 2018.

- Stearse, Thomas, et al. "The Association between Academic Pressure and Adolescent Mental Health Problems: A Systematic Review." *Journal of Affective Disorders*, vol. 339, Oct. 2023, pp. 302–17. ScienceDirect, <https://doi.org/10.1016/j.jad.2023.07.028>.
- Streit, Cara, et al. "Relations Among Parenting, Culture, and Prosocial Behaviors in U.S. Mexican Youth: An Integrative Socialization Approach." *Child Development*, vol. 92, no. 4, July 2021. Crossref, <https://doi.org/10.1111/cdev.13550>.
- Tang, Xinfeng, et al. "Psychosocial Risk Factors Associated with Depressive Symptoms among Adolescents in Secondary Schools in Mainland China: A Systematic Review and Meta-Analysis." *Journal of Affective Disorders*, vol. 263, Feb. 2020, pp. 155–65. ScienceDirect, <https://doi.org/10.1016/j.jad.2019.11.118>.
- Thomson, Kimberly C., et al. "Parent Psychological Distress and Parent-Child Relationships Two Years into the COVID-19 Pandemic: Results from a Canadian Cross-Sectional Study." *PloS One*, vol. 18, no. 10, 2023, p. e0292670. PubMed, <https://doi.org/10.1371/journal.pone.0292670>.
- Tung, Keith T. S., et al. "Family Financial Pressure in Childhood and Telomere Length in Early Adolescence: A Prospective Study." *Genes*, vol. 13, no. 5, Apr. 2022, p. 721. PubMed, <https://doi.org/10.3390/genes13050721>.
- Xiao, Sonya Xinyue, et al. "Family Economic Pressure and Early Adolescents' Prosocial Behavior: The Importance of Considering the Types of Prosocial Behavior." *Journal of Adolescence*, vol. 95, no. 6, Aug. 2023, pp. 1220–31. Crossref, <https://doi.org/10.1002/jad.12196>.
- Yan, Jia, et al. "Longitudinal Associations between Parent-Child Relationships in Middle Childhood and Child-Perceived Loneliness." *Journal of Family Psychology [US]*, vol. 32, no. 6, 2018, pp. 841–47. APA PsycNet, <https://doi.org/10.1037/fam0000446>.
- Yang, Qi, et al. "The Association between Perceived Family Financial Stress and Adolescent Suicide Ideation: A Moderated Mediation Model." *Behavioral Sciences (Basel, Switzerland)*, vol. 13, no. 11, Nov. 2023, p. 948. PubMed, <https://doi.org/10.3390/bs13110948>.

Redefining Richard Florida's Creative Class (CC): Artificial Intelligence and the Future of the Creative & Cultural Industries (CCIs) By Peri Yeung

Abstract

The evolving landscape of the Cultural and Creative Industries (CCIs) amid the rapid integration of Artificial Intelligence (AI) in the sector is raising questions regarding authenticity, creativity, authorship, and the future of creative work. This study explores how creative professionals working in the CCIs perceive the impact of AI on their work and occupational identity. Simultaneously, Richard Florida's influential concept of the Creative Class highlights the centrality of creative professionals in driving growth, but the rapid adoption of AI necessitates revisiting this theory. An anonymous online survey was employed to capture both qualitative and quantitative insights into AI adoption, associated benefits and risks, and ethical considerations within creative workflows. This study contributes to the critical understanding of how technological changes reshape creative labour and the perceptions of creative professionals. Therefore, this necessitates a need to revisit the concepts and definition of creativity, as well as policy frameworks that are relevant to cultural production.

1. Introduction

The Cultural and Creative Industries (CCIs) serve as vital economic engines, moving from the periphery of economic engines to the very epicentre of strategies that nations, countries, and cities have adopted globally. These industries protect our cultural richness, preserving the essence of human creativity and intangible aspects of life such as well-being and identity. Consequently, the CCI's have established a domain in which the most important role is human creativity; but is this truly the case? The rapid development of Artificial Intelligence (AI) within the past few years is a paradigm shift for the CCIs, offering a multitude of advantages and challenges. The idea that the creative sector is a domain in which the core role material is human creativity will be challenged further on in the Literature review.

Meanwhile, Richard Florida's thesis on the Creative Class has long dominated discussions about the spatial organisation and economically mobilisation of creativity. The Creative Class represents a distinct group of individuals who play a pivotal role in driving regional economic growth through their innovative and creative endeavors. This group is composed of individuals that engage in work involving creating meaningful new forms, ranging from scientists, engineers, artists, writers, designers, architects, to knowledge-based professionals in sectors like high-tech industries, finance, and business management. But what happens when AI, a non-human influence, is introduced to the Creative Class and the CCIs?

AI's integration promises to revolutionise creative processes by enabling firms to allocate more resources towards other essential aspects of creativity, such as ideation. In other words, this frees human creators from mundane tasks, allowing them to focus on more conceptual or imaginative work, augmenting human creativity. This also makes the creative process more efficient. However, the increasingly common use of AI in the CCIs also raises concerns

regarding authenticity within the creative industries, and the preservation of genuine human creativity in the face of technological influence (Lee, 2022). As questions surrounding authenticity and ethics about the usage of AI in the CCIs surface, issues of ownership, bias, and intellectual property rights emerge as significant considerations. Moreover, AI greatly disrupts the spatial logic of creativity, which is central to Florida's theory. Where physical urban hubs once attracted creative talent, virtual platforms and datasets now serve as the new creative capitals that are accessible globally. This shift removes the link between creativity concentration to geographical constraints, redistributing opportunities. However, this also risks exacerbating inequalities, as access to AI tools and training becomes evermore significant. When one thinks of creativity, it is usually solely attributed to human ingenuity and imagination. However, with the integration of AI, the conventional notions of creativity are challenged. This prompts a reevaluation of creativity within the CCIs and the Creative Class' function within this increasingly technology-driven system that emerges.

The involvement of AI in the creative sphere has deeply altered the processes and economies of cultural production, revolutionising various creative fields by assisting with their respective creative processes such as producing music, writing books, and generating parts or all of a film. The increasing reliance on AI algorithms and predictive models fundamentally alters and influences how cultural products are conceived, produced, experienced and consumed. This is seen commonly in apps asking for permission to track activity across other apps and websites in order to personalise advertisements for the user. Thus, the roles and identities of creative labour within the CCIs are reshaped, as AI continues to influence the processes that guide their work and redefine the standards and values assigned to their creations. This transformation challenges the boundaries and notions that have historically defined the CCIs and the creative class. The presence of AI begs the question: What does it mean to be a cultural or creative worker and innovator in a landscape where artificial intelligence plays a significant role in the creative process? The ever-increasing use of AI in the creative process blurs the boundary between human and machine creativity, leading to implications of a future in which no human participates in a creative process, and human input is no longer central to the creative process, thus raising questions about the evolving nature of creativity and the shifting dynamics of artistic expression in this digital age.

The paradigm shift caused by AI prompts a critical reexamination of the traditional structures and hierarchies within CCIs, necessitating a reevaluation of how creativity is defined and valued within the creative industries where AI increasingly shapes the cultural landscape. This research aims to examine what it really means to be a creative worker, in an era where the lines between human and machine creativity are increasingly blurred; it will attempt to capture how creative professionals perceive and integrate AI in their creative work, paying particular attention to changes in authorship, ownership and occupational identity. What new skill sets are emerging? And, most importantly, how do these changes redefine Richard Florida's Creative Class?

This study employs a pragmatic mixed-method design, combining quantitative survey data with qualitative insights to examine how AI is reshaping the creative workforce. Data will be collected through an anonymous online survey targeting five key stakeholder groups, namely: Creative Professionals (artists, musicians, writers, designers), Cultural Industry Professionals (curators, arts administrators), Technology Experts (AI researchers, data scientists), Urban Planners, Developers, Academics, and Entrepreneurs in creative fields. The survey is structured to assess the current AI adoption in creative workflows, perceived benefits and risks of AI, and ethical and philosophical debates around AI-generated content.

2. Literature review

2.1 The CCIs

Over the last 20 years, there has been a growing recognition of the vital role played by CCIs in stimulating regional development, social innovation, and cultural vitality. They are now seen as essential economic engines that combine business, creativity, and culture to create jobs, stimulate commerce, and revitalise cities. In personal terms, Cultural and Creative Industries are characterised as sectors rich in innovation, responsible for generating creative outputs that not only influence society but also have an intangible cultural value and heritage significance. This is especially important as heritage is the essence of our cultural identity, encompassing traditions, artifacts, and values passed down through generations. It serves as a link between the past, present, and future, preserving humanity's shared history. The Department for Digital, Cultural, Media and Sport's (DCMS) definition of the Cultural and Creative Industries (CCIs) outlines them as "those industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property" (DCMS, 2001). Meanwhile, UNESCO defines them as "sectors of organised activity whose principal purpose is the production or reproduction, promotion, distribution and/or commercialisation of goods, services and activities of a cultural, artistic or heritage-related nature" (UNESCO, 2021).

The importance of the cultural and creative industries (CCIs) lies in the fact that they are critical to the economy. These industries, having the capacity for wealth creation and job generation through the creation and utilisation of intellectual property, are significant economic drivers and contribute greatly to the world's GDP, while also providing employment for many (DCMS, 2001). However, besides being solely an economic driver, the cultural and creative industries are also one of the few sectors that provide cultural richness; it is simultaneously important to preserve the human elements of creativity and the intangible aspects of life such as life and well-being. These sectors fuel innovation by generating new products, ideas, and approaches, fostering cross-industry innovation and a more innovative workforce through arts education. Moreover, the CCIs are recognised for their creative capacity in engaging audiences in co-creation processes, particularly in fields like video games, music, and design. Post-COVID, the CCIs have embraced technology more than ever and have been at the forefront of

experimenting with innovative digital production and distribution models, generating substantial revenue and driving cultural participation (OECD, 2021).

Prior to the widespread emergence of Artificial Intelligence (AI), the nature of creative work was already undergoing significant changes due to various factors such as user-generated content, digitalised media, and technological advancements. User-generated content consisted of individuals utilising platforms such as social media, blogs and forums to express themselves creatively, creating and sharing content with a global audience. Through these platforms, creators would receive direct-feedback from their audiences in real-time, leading to more responsive and audience-centric content. Digitalisation also made creative works more accessible and easier to distribute globally. Creative individuals could reach a wider audience without the limitations of physical mediums. Furthermore, with the advancement of technology, individuals now have access to more resources and tools. For example, a novice could learn the skills parallel to that of a graduate-level art course through online tutorials. Similarly, advancements in creative tools and software also democratised access to professional-grade resources, allowing more people to engage in creative pursuits. Technologies like design software, video editing tools, and music production software also increased efficiency in creative workflows, augmenting human creativity by redirecting the focus to the process of ideation. Most recently, Augmented Reality (AR) and Virtual Reality (VR) technologies opened up new frontiers for immersive storytelling, gaming, and interactive experiences, contributing to the ever-blurring boundaries between physical and digital worlds.

2.2 The Creative Class

The Creative Class, as conceptualised by Richard Florida, refers to a distinct group of individuals who play a pivotal role in driving regional economic growth through their innovative and creative endeavors. Florida's theory accentuates the vital role of creative workers in driving economic growth, urban development, and innovation. Individuals of the creative class are crucial for fuelling regional economic growth through their creation of novel forms that elicit innovation. The theory also recognises creative professionals across various sectors like high-tech industries and healthcare as pivotal contributors to economic advancement, highlighting the significance of creative individuals in driving problem-solving and innovation. What unites these professionals is their reliance on creativity. The common characteristics across creative classes consists of flexibility, self-expression, and a demand for autonomy. These attributes are crucial in distinguishing human creativity from machine-generated output, especially in this era of AI.

The ethos and values of the creative class values experiences that are active, authentic, and self-structured. Their emphasised appreciation of creative productivity is also particularly relevant in the context of AI-generated art, simultaneously highlighting the distinction between human and machine creativity. Furthermore, the creative class prefers hands-on activities such as rock climbing, cycling, and travel to remote locations that challenge them physically and intellectually; this preference for active engagement over a passive spectatorship is a defining

trait of the Creative Class lifestyle. Consequently, this places a significance on street-level culture in fostering creativity and community engagement. Indigenous street scenes, filled with small venues, art galleries, music clubs, and diverse cultural offerings, serve as hubs for creative stimulation and social interaction. These organic and dynamic spaces provide opportunities for individuals to immerse themselves in diverse experiences and connect with other artists and creators. However, this also gives rise to the commercialisation of experiences, such as commodifying cultural events, leading to an inauthentic, packaged experience that lacks originality and depth. Creative individuals should instead seek out genuine, indigenous cultural offerings that resonate with their creative sensibilities (Florida, 2012).

The emergence of "creative centres," characterised by high concentrations of creative workers, gives rise to hubs that foster economic success and growth, attracting the creative class seeking validation and diverse experiences. This theory also highlights the migratory trends of creative workers towards these creative centres, emphasising their role in shaping urban growth. Regions with a notable concentration of the creative class often excel in innovation and high-tech industries; this demonstrates the critical link between a region's creative workforce and its innovation potential (Florida, 2003). To draw in the Creative Class, the city must have what Florida calls the "three T's". Talent, which refers to an exceptionally skilled and educated population; Tolerance, which refers to encompassing a diverse community with a culture of mutual respect and acceptance of differing viewpoints; and Technology, which refers to the technological infrastructure essential for fostering an entrepreneurial climate.

In addition to "creative centres," bohemian amenities also play a crucial role in nurturing the Creative Class. Bohemian amenities include a range of cultural and lifestyle elements that reflect the bohemian ethos and values, such as prioritising pleasure and happiness in the form of experiencing and appreciating life's offerings. The bohemian lifestyle emphasises the pursuit of joy and fulfilment through artistic expressions, cultural engagements, and personal experiences to enrich an individual's life. Individuals who embrace bohemian values often exhibit a deep respect for aesthetics and find meaning and inspiration in the artistic and creative aspects of life. The spiritual and sociopolitical dimensions of bohemian amenities are also significant. These dimensions tend to be more intuitive rather than logical, focusing on individualism over conformity and encouraging unique perspectives on societal issues and personal beliefs. Bohemian culture often challenges traditional norms and values, fostering a spirit of free thought, creativity, and nonconformity in both personal and collective expressions.

However, there are also critiques of the theory of the creative class. Some challenge Richard Florida's advocacy for attracting the "creative class" in urban planning and economic development, questioning the alignment of his strategy with the preferences of many creative individuals. It criticises Florida's over-emphasis on creative workers enjoying the urban lifestyle. Instead of Florida's accentuation on fostering bohemian, high-density areas, many creatives actually prefer suburban living with amenities like easy commutes, safe environments, good schools, and low taxes. Empirical analyses suggest that bohemian factors have limited impact on urban success compared to education levels, emphasising the importance of catering to skilled

workers' needs for sustainable urban development over creating trendy urban environments (Glaeser, 2005). Another point to be raised is that while creativity, skills, and diversity are vital in urban development, the theory of the Creative Class may oversimplify the multifaceted nature of city success. Florida's approach attempts to combine different factors like cultural diversity, talent, and economic performance but falls short in capturing the full complexity of urban development. It underscores the importance of a holistic view that includes physical, economic, social, and cultural aspects in shaping cities. Additionally, while creativity may have a positive impact on prosperity, more emphasis is placed on skills, talent, and ethnic diversity in driving urban growth and innovation (Nathan, 2008). Most strikingly of all, critics point out discrepancies between Florida's idealised portrayal of the creative class as affluent and thriving and the harsh realities faced by many individuals in creative fields, particularly in cities like Vancouver. Florida's narrow depiction of the creative class as predominantly residing in affluent neighbourhoods is challenged, highlighting the diversity of creative practitioners found in lower-income areas. Moreover, Florida's simplistic definition of "creative" is called into question, emphasising that the creative class encompasses individuals ranging from struggling artists to successful CEOs, thus debunking the notion that all members enjoy high incomes. The widening economic disparities amplified by the creative economy highlights the growing gap between the creative class and the service class, leading to concerns about economic segregation and the marginalisation of certain groups, notably in costly urban environments. Furthermore, it is ambiguous whether the concept of the "creative economy" mentioned fosters a form of "hyper-capitalism", underscoring the importance for collective action and social cohesion within the creative community to address socio-economic inequalities and advocate for improved working conditions (Dembicki, 2014).

2.3 Impact of AI on the creative industries

AI has recently gained traction in the CCIs as it becomes increasingly integrated into those industries; this has been driven by an ongoing breakthrough in data analytics, accompanied by AI platforms becoming more affordable and accessible. AI refers to the simulation of human intelligence processes by machines, typically computer systems. These processes include learning, reasoning, problem-solving, perception, and decision-making (Saini, 2023). AI includes a wide spectrum of technologies that aim to mimic cognitive functions such as learning from data, adapting to new situations, and performing tasks that typically require human intelligence. Historically, the act of creation in the CCIs have relied on human intuition; refining ideas, absorbing criticism and taking in collaborative feedback is what enabled creatives to explore concepts and creations that they might otherwise might not have made. However, how has the integration of AI fundamentally changed the creative process?

Over the years, the evolution of AI has progressed from subtle applications, such as AI providing recommendations and engaging in chat interactions through bots, to more advanced stages where artists are capable of exclusively generating art using AI technologies. Early implications of AI in the creative sector include being used in various aspects of film production,

including scriptwriting, editing, and special effects generation (Gao and Leong, 2024). Tools like IBM's Watson have been used to analyse scripts and predict audience reactions to optimise storytelling. AI-driven design tools like Adobe's Sensei and Canva's Design School also leverage machine learning to assist users in creating visually appealing graphics and layouts.

Nowadays, an integral part of AI's involvement in the CCIs is its generative deep neural network-based artwork. Contrary to common conceptions, the definition of "AI art" does not solely refer to the output created by artificially intelligent systems. Instead, it encompasses the utilisation of algorithms and intelligent techniques within the artistic process. This perspective diverges from traditional views, emphasising that AI itself is a medium for artistic expression, regardless of the level of autonomy within the system. In this context, AI art is not just about the end result but the utilisation of intelligent technologies in the creation process, expanding artistic boundaries and possibilities beyond the constraints of human capabilities (Ruth & Andrés, 2020). This gives rise to the term "computational creativity", which encompasses the intersection of AI and creativity. An example of this is the music style and harmony transfer technology developed at the SONY Computer Science Lab. This technology assists composers in harmonising music pieces across different genres, showcasing the innovative potential of AI in creative endeavors (Mántaras, 2017).

In today's rapidly digitalising world, the emergence of Artificial Intelligence (AI) poses a plethora of benefits to the CCIs. The integration of AI can enhance the creative process by increasing the efficiency of creative labour and decreasing the costs of production in fields like filmmaking and creative writing (Esling & Devis, 2020). This allows firms to reallocate resources towards more crucial parts of the creative process, such as ideation. AI can also augment human creativity, automating mundane tasks to allow more people to express themselves in more engaging tasks, thus making people more creative in their work. Furthermore, AI is capable of bulk-generating creative content like movies and books using computational creativity, taking advantage of economies of scale and increasing the volume of output.

As AI continues to automate more and more mundane tasks and analyse large amounts of data such as music in order to create new music, challenges start to arise. For instance, the potential job displacement as a consequence of AI is a prominent concern as it highly threatens traditional roles within CCIs, leading to an ever-increasing gap between those adept at AI integration and those left behind. The future of genuine human creativity amidst AI's influence is also uncertain (Mahmud et al., 2023). As AI evolves and progresses, its imitation of human work becomes ever more similar, having the ability to rival human creativity, such as human art being hard to differentiate from AI-generated art. AI currently lacks the necessary causal connections to exhibit true intentionality, as highlighted by Searle's Chinese Room argument. It poses a challenge to the notion of creativity in AI, emphasising the importance of understanding the underlying mechanisms that drive creative processes. However, the potential for future advancements in AI, particularly in the development of more embodied artificial intelligences equipped with advanced sensors for environmental interaction, offers a glimpse into a future

where AI may possess the causal connections required to imbue meaning and intentionality into creative endeavors (Mántaras, 2017). This raises the question of whether or not AI will override human creativity, and if AI itself can be considered creative. Finally, an overreliance on AI may also stifle real and genuine human creativity (De Cremer et al., 2023). It may be a challenge for the creative industries to become less creative as processes within these industries are increasingly automated with the use of AI.

Questions of authenticity are also raised (Lee, 2022); there are many contrasting opinions about the ethical and legal implications associated with AI in CCIs. Ownership and authorship makes up the notion of authenticity. AI is a predictive tool based on algorithms, and as such, it plays a significant role in decision-making processes, from approving job applications to generating creative content. However, the inherent biases within these algorithms can lead to discriminatory outcomes, favoring certain groups while disadvantaging others (Small, 2023). For instance, AI systems may exhibit biases against minority ethnic groups or women due to the way they are programmed, perpetuating systemic inequalities. Additionally, the question of ownership and intellectual property rights surrounding AI-generated content remains a complex issue. The discussion of ownership of AI-generated content is a blur. As AI continues to evolve and contribute to content creation, concerns about algorithmic bias and the rightful ownership of AI-generated work come forward, challenging traditional notions of authorship and creativity in this digital age.

2.4 Impact of AI on the Creative Class

In exploring the intersection of Richard Florida's framework of the Creative Class and the evolving landscape of AI creativity, it is imperative to analyse how these concepts align or clash with the realities of AI. Florida's conceptualisation of the Creative Class emphasises the pivotal role of creative individuals in driving economic growth and innovation. Central to his theory is the idea of a "super creative core," composed of highly skilled members who propel advancements in various sectors. Florida (2003) states that technology is central to attracting creative talent, praising technology as a key component of urban development. However, this was envisioned in a time when AI was still largely science fiction. The value Florida places on technology becomes increasingly relevant in this era of AI, where highly skilled creative workers proficient in AI tools can potentially emerge as 'super creatives.' By leveraging computational creativity resources to produce art, technology becomes a catalyst for these individuals as they experience enhanced mobility within and across the creative group.

In contrast, as technology, particularly AI, advances, this may introduce complexities within the creative sector. The widespread adoption of AI has the capability to replicate activities traditionally exclusive to human ingenuity, such as music composition, filmmaking and painting, which may actually limit the dynamics of the sector. The blurred distinction between human and AI-generated creativity poses a threat to creative jobs, potentially diminishing the value of the 'super creative core' as AI becomes more prevalent. As the usage of AI becomes ever more prominent, it may also stifle human creativity. Thus, creative workers may feel threatened as

unskilled workers are increasingly capable of taking over their jobs. This paradigm shift could lead to a decrease in the number of 'super creatives' and create challenges for creative workers who are unable or unwilling to adapt to the changing landscape. What happens now if humans are no longer needed?

Moreover, Florida's spatial logic of creative clusters also needs to be re-evaluated in light of AI. His "3 T's", Talent, Tolerance, and Technology, are made under the assumption that physical proximity in urban hubs fosters innovation. However, with the emergence of AI, the "3 T's" need to be reexamined. Pre-AI, Talent was tied to place-specific skills and networks; now, it encompasses hybrid-techno-creativity—a fusion of technical AI literacy (e.g., prompt engineering, algorithmic collaboration) and traditional creative skills. Platforms like OpenAI's Codex democratise coding talent, enabling distributed teams to innovate beyond geographic limits. This shift also fosters meta-creativity: the ability to curate, remix, and optimise AI-generated outputs, which becomes a new form of cultural capital. However, not everyone has equal access to these tools, creating new divides. Florida's assumption that place attracts talent is obsolete when creative labour can occur in virtual spaces, however, this democratisation is uneven, privileging those with AI access.

Next, Tolerance. The Creative Class celebrates diversity as an economic asset, but AI disrupts this premise. While creative workers traditionally thrived on cultural pluralism, AI now acts as the primary agent of 'inclusivity', or, in other words, exclusion. AI discriminates innately by design (Heinrichs, 2022). Its epistemic opacity and training data biases (e.g., Western-dominated datasets) homogenise creative outputs, contradicting Florida's vision of tolerance. Platforms like TikTok exemplify how algorithmic curation connects creators globally while reinforcing systemic biases; algorithmic "tolerance" is conditional, favoring content that aligns with dominant cultural norms. This risks marginalising voices from the Global South and other underrepresented groups, undermining the Creative Class's reliance on diversity as an engine of innovation.

Finally, Florida's Technology pillar assumes urban density accelerates innovation through shared infrastructure. AI transforms this relationship: rather than being a supporting tool, AI becomes a co-creative agent. For the "super-creative core," AI is no longer just a pipeline but a collaborator, generating designs, music, or code. Yet this partnership comes at a cost. AI also acts as a gatekeeper; those without access to expensive tools (e.g., generative AI subscriptions, high-end hardware) are excluded, regardless of their innate creativity. The fact that intangible capital increasingly drives creative economies (Hearn, 2020) underscores this shift: if the means of production are digital, urban density matters less, but control over technology matters more. This creates a paradox in which AI expands creative possibilities, while simultaneously consolidating power in the hands of those who own or monetise these tools.

The same tools that enable creators to connect and collaborate across the globe also enable a new form of digital exploitation where said creators' content is trapped in corporate platforms that profit from their labour without compensating them (Collie & Wilson-Barnao, 2024). Florida's spatial theory is partially obsolete because AI decouples creativity from place,

but it fails to account for the digital enclosures that replace geographic barriers. Talent is no longer bound to cities but is filtered by AI access, Tolerance is mediated by opaque algorithms, and Technology's role shifts from enabler to gatekeeper.

AI's profound influence on urban planning and creative endeavors may reshape the very composition of the creative class. The increased reliance on high-end technological infrastructure could usher in a new form of cultural capital, raising questions about the necessity of human involvement in creative processes when AI systems can also produce art. As AI disrupts traditional notions of creativity, this prompts a reevaluation of creative labels and work, forcing a critical examination of what constitutes creativity in the presence of AI. The evolving role of AI in the creative industries may necessitate a reexamination of the creative class structure and the dynamics that shape the sector, such as the creation of an AI-dependent Creative Class whose power is entwined with the ability to use AI platforms, where inequalities emerge from differing levels of technical fluency, such as the financial barriers to using AI, such as firms purchasing services of costly AI platforms, in order to gain access to certain datasets. The potential transformation of the workforce, driven by AI's capabilities, coupled with the transformations undertaken in regard to the "3 T's" could redefine the essence of creativity and the future of creative professions, stretching Florida's existing framework of the Creative Class.

2.5 Bridging gaps in literature surrounding the intersection of AI, CCI's, and CC

The intersection between the AI, CCIs, and the Creative Class has been extensively studied, yet critical gaps remain in understanding how AI-mediated cultural production truly reshapes these frameworks. Moreover, the rapid adoption of AI tools has fundamentally altered the ontology of creative work, yet the actual experiences of creative workers navigating this shift are barely touched upon. Already-existing theories, including Florida's Creative Class, written in 2002 and revised in 2012, were formulated in a pre-AI era, where the capabilities of AI today only existed in the likes of science fiction back then. This time gap leaves contemporary dynamics, such as AI's role in creative workflows and worker perceptions, underexplored. As such, it would be beneficial to update those theories. Addressing these gaps in literature is especially vital for educationalists designing curricula (e.g., generative AI modules for art schools), and creative professionals navigating AI's disruptive potential (Hearn, 2020; Manovich, 2019).

For instance, propositions that AI art should be understood as a medium rather than an autonomous creator challenges the traditional notions of authorship, but this theoretical stance has yet to be backed up with real, observed studies on how artists perceive their agency when co-creating with algorithms (Manovich, 2019). Similarly, the Creative Class puts technology as an enabler of creative mobility (Florida, 2002), yet the framework did not anticipate AI's dual role as both a democratising force and a disruptor of creativity. The democratisation argument, that AI lowers barriers to entry and thus allows novices to produce professional-grade work through tools like MidJourney or ChatGPT, threatens to devalue the skills of creative workers,

potentially destabilising the "super-creative core" that Florida deemed essential for urban innovation.

Moverover, there is a conflict between arguments of efficiency versus authenticity in AI-mediated cultural production. On one hand, firms can make use of AI to streamline workflows, as seen in robo-journalism's automated news generation (Hearn, 2020), or algorithmic design optimisation in architecture (Stewart et al., 2024). These applications align with Florida's emphasis on technology as a productivity multiplier, yet they also risk reducing creativity to that of a mere commercialised good, where algorithmic predictability supplants human individualism.

The muddled ethical boundaries of this shift into the era of AI is also prominent. When AI systems trained on copyrighted datasets produce derivative works, who does the work generated truly belong to? Naturally, questions of intellectual property and labour exploitation will start to surface. For example, the rise of AI-generated music platforms like Boomy has sparked debates about whether algorithmic compositions dilute the cultural value of music as an expressive form, or simply prolong the history of technological mediation in art. Such debates underscore the need to redefine creativity: is it the output that matters, or the intentionality behind it? Searle's Chinese Room argument, as cited in critiques of computational creativity (Mántaras, 2017), suggests AI lacks intrinsic understanding, yet its outputs increasingly mimic human creativity, blurring the line between human and AI creativity.

Ultimately, the unresolved question is whether AI augments or annexes human creativity. While some believe that AI will expand artistic possibilities, others warn of a "post-creative" future, where human input is reduced to prompt engineering (De Cremer et al., 2023). The Creative Class's resilience depends on how well creative workers can adapt their skill sets and mindset. For example, artists fluent in using AI may emerge as the new "Super-creative core" (Florida, 2002), while those reliant on more traditional techniques may face displacement. This parallels broader labour trends that have been observed under the capitalism of AI, in which those who are capable of harnessing AI have a higher value than those who resist automation, which brings me to the point of this research – that is, to examine the very definition of being a creative worker, in this era where AI is so prominent. There is a need for the framework of the Creative Class to be reworked with the new realities of AI. Without this connection, theories of the Creative Class risk becoming artifacts of a pre-algorithmic age, ill-equipped to explain the shifts underway in the CCIs.

3. Methodology

This study adopts a quantitative survey-based research design to examine creative professionals' perceptions of artificial intelligence (AI), its adoption within their work, and the implications for occupational identity within the Cultural and Creative Industries (CCIs). Surveys are a quantitative instrument, providing breadth of knowledge, meaning that it does not allow for the opportunity to gain in-depth knowledge about the participants. The decision to use an online survey stems from its strengths in generating breadth of knowledge across a relatively

large sample, standardising responses for comparative analysis, and enabling cost-effective data collection across diverse geographical contexts.

Consequently, an online questionnaire was employed to capture both quantitative and qualitative data. The survey consisted of 20 quantitative questions (these included 5-point Likert scale and multiple-choice items) and 3 open-ended qualitative questions, which were optional. The survey was divided into five sections:

- 1) Agreement to Participate: Participants were assured of voluntary participation and informed consent
- 2) Demographic Information: Participants were required to provide their age, gender, continent of residence and professional field within the CCIs
- 3) Familiarity with AI: A section regarding participants' knowledge and experience of AI tools in their creative practice
- 4) The Future Outlook of AI: Gauging participants' perceptions of AI's impact on creativity, employment, authenticity and occupational identity
- 5) Additional comments: An optional, open-ended section to allow participants to provide reflections on AI's role in creative work

Additionally, the survey was piloted with three creative workers to refine clarity and relevance of the questions before wider dissemination. Responses from pilot tests were not accounted for in the final results.

The collected quantitative data were analysed using a mixed-methods approach combining descriptive, inferential. Descriptive statistics such as frequencies, means, standard deviations were used to characterise the sample demographics and provide a univariate summary of key variables, such as the distribution of professions within the creative sector and general attitudes towards AI. To examine relationships between variables, inferential statistical techniques were applied. Given the ordinal nature of Likert-scale data, non-parametric tests were prioritised:

- Chi-square tests of independence were used to examine associations between categorical variables, such as professional domain (e.g., visual arts, music, writing) and nominal attitudes (e.g., enthusiast, sceptic).
- Kruskal-Wallis H tests were employed to compare the means of scores across more than two professional groups.
- Ordinal regression analysis was conducted to model relationships between ordinal dependent variables (e.g., frequency of AI usage) and independent predictor variables (e.g., age, profession, perceived benefits), allowing for the identification of significant factors that influence AI engagement.

All analyses were conducted using Jamovi, an R-based open-source statistical analysis software. Qualitative data from the three open-ended questions were analysed through basic thematic coding.

Ethical conventions of academic research have been upheld; by default, the survey's design is anonymous and all data is confidential, only being used for the purposes of this research. The survey was employed through Google Forms, a platform that ensures the anonymity of respondents. Informed consent was ensured as all participants were required to click 'start' to begin the questionnaire, and voluntary participation was clearly stated as participants were allowed to quit at any time.

4. Results and discussion

This section weaves together the findings from three layers of analysis, non-parametric group comparisons, a chi-square test of independence, and an ordinal–logistic regression. At its core, this research aims to examine how the adoption of AI reshapes the views, perceptions, concerns and identities of creative workers and how this informs Florida's Creative Class theory.

4.1 Results

4.1.1 Experience bands show little attitudinal divergence

A Kruskal-Wallis H test revealed a marginally insignificant difference ($\chi^2(3) = 7.21$, $p = 0.065$) in perceptions of AI's impact based on years of experience in creative fields, with a moderate effect size ($\varepsilon^2 = 0.1414$). Post-hoc analysis using Dwass-Steel-Critchlow-Fligner (DSCF) pairwise comparisons showed that respondents with 8–15+ years of expertise were significantly more likely to hold negative views ($W = -2.064$). This suggests that while novices perceive AI as a tool to bridge skill gaps, for example, enabling non-musicians to generate songs. Therefore, experienced professionals may see AI as a threat to their creative autonomy and occupational identity. This shows a characteristic of the AI skeptic, in which they are usually workers with more experience in the field.

4.1.2 Gender derives a significant divergence in the importance of creativity

Despite an initially insignificant chi-square result, the Kruskal-Wallis H test detected a statistically significant gender-based difference in attitudes toward the essentiality of human creativity ($\chi^2 = 4.98$, $p = 0.026$, $\varepsilon^2 = 0.0977$). Women (27 out of 52 respondents) were more likely to rate the importance of human creativity lower (modal response = 3) compared to men (modal response = 4 or 5). This aligns with broader research suggesting that women are more skeptical of AI due to ethical concerns, such as bias in training data or questions about ownership of AI-generated content.

4.1.3 Education level predicts the type of AI concern

Contingency tables show that someone who is highly educated is more likely to be concerned with ethical implications. A chi-squared test reveals a statistically significant result, $\chi^2(28) = 44.6$, $p = 0.024$. Post-hoc tests using Pearson residuals show that Doctorate holders are most concerned regarding the ethics of using stolen and artificial content with a value of 4.903, Master's holders are most concerned with the loss of authenticity (0.812), while Bachelor's degree holders are mainly concerned regarding job displacement (0.7933).

4.1.4 Field of expertise marginally impacts perceived authenticity

Though a Kruskal–Wallis H test revealed a marginally statistically insignificant difference between the respondent's primary creative field of expertise and respondents' perceptions about whether content created entirely by AI can be considered “Authentic creativity”, $\chi^2(17) = 25.8$, $p = 0.078$, with a moderate effect size ($\epsilon^2 = 0.506$). Post-hoc. According to the Dwass-Steel-Critchlow-Fligner pairwise comparisons, coders are more likely to regard content created entirely by AI as “authentic creativity” than designers (e.g. fashion designers, product designers, graphic designers, etc.), with a W-value of 2.963. In other words, creative fields that require creativity in relation to the visual arts feel less positively about calling AI-generated content “authentic creativity”.

4.1.5 Usage frequency drives attitude polarisation

A Kruskal-Wallis H test revealed a significant difference ($\chi^2(4)=24.39$, $p<0.001$) between respondents' usage frequency and how they perceived the overall impact of AI on their work, with a moderate effect size ($\epsilon^2=0.478$). Post-hoc analysis using DSCF pairwise comparisons showed that heavy users are more inclined to perceive the overall impact of ai on their creative processes as positive, whereas sporadic users are more inclined to perceive the overall impact of ai as negative ($W = -4.551$). All the differences concentrate at the extremes of frequency of AI usage; the more one uses AI, the more legitimate it is. Moreover, frequent users of AI are more likely to consider content created entirely by AI as “authentic creativity. A Kruskal-Wallis H test revealed a significant difference ($\chi^2(4)=9.73$, $p=0.045$), with a moderate effect size ($\epsilon^2=0.191$).

4.1.6 Multivariate analysis confirms frequency effect and ‘Master’s’ premium

An ordinal-logistic regression modeled the same “overall impact” item with three predictors: AI-use frequency (dummy coded, reference = “Always”), education level (reference = “Bachelor’s”), and stance on AI-generated ownership. The model explains 31 % of deviance (McFadden $R^2 = .308$) and reveals that frequency is the dominant driver and a strong predictor in determining one's perception of AI. Meanwhile, education only matters selectively as masters degree holders are shown to foster a more optimistic view of AI, and ownership beliefs add no explanatory power once use-frequency and education are controlled ($p > .30$ for all contrasts).

The more frequently people work with AI, the more favourable their overall appraisal, until the saturation at “Often”. Heavy users and “often” users report similar findings; the big gap

is between frequent users and everyone who uses AI sporadically or not at all (“Never” users). In the survey, only respondents with a Master’s degree stand out from this group as statistically significantly more positive than a Bachelor’s degree with an odds ratio of 5.98827; in other words, philosophising about IP ownership does not explain how positively people feel about AI’s impact.

4.2 Discussion

4.2.1 CCIs/Creative workers

The most critical determinant of AI acceptance is frequency of use. Frequency of AI engagement dwarfs professional seniority, legitimising AI through use as those who engage with AI tools regularly are more likely to view it in a beneficial light; frequent users are also more likely to see AI’s outputs as authentically creative. This aligns with technology adoption theories, where exposure reduces resistance over time (Liao et al., 2009). However, creative workers cannot be treated as a monolithic group: while statistical significance was marginal, trends indicate a disciplinary divide, with coders embracing AI’s role in creativity more readily than designers. This reflects differing values: coders prioritise functionality and problem-solving, whereas designers emphasise human craftsmanship and originality. Additionally, experience level shapes attitudes, with veterans (8+ years in their field) more resistant to AI, likely due to fears of devaluation. Meanwhile, gender plays a role in ethical skepticism, with women more cautious about AI’s implications for creativity. These findings suggest that AI’s integration into creative work is not just a technical shift but a cultural and ideological one, requiring nuanced approaches to training and policy.

4.2.2 Implications for Florida’s Creative Class Theory

Richard Florida’s Creative Class theory, originally framed around the “3T’s” (Technology, Talent, Tolerance), requires substantial revision to account for AI’s transformative impact on the Creative Class. In Florida’s original framework, technology served as an enabler of creativity, a means to streamline production or enhance artistic expression. However, AI’s role has expanded beyond facilitation; it now co-creates alongside humans, generating text, music, and visual art autonomously. This shift introduces new tensions, particularly around authorship and job displacement. For instance, generative AI can now produce marketing copy or album artwork, roles traditionally filled by human creatives. Consequently, technology is no longer just a catalyst for creative work but also a gatekeeper, determining which skills and workers remain relevant.

Meanwhile, Florida’s definition of “talent” emphasised innate creativity and specialised skills. In this era of AI, however, talent is increasingly defined by orchestration—the ability to effectively guide, refine, and collaborate with AI tools. For example, a graphic designer’s value may no longer hinge solely on their ability to draw but on their proficiency in prompt engineering, model fine-tuning, and curating AI-generated outputs. This redefinition disrupts

traditional creative hierarchies, as newcomers with strong technical-AI hybrid skills may outperform veterans resistant to adaptation.

The third T, Tolerance, was originally linked to diverse and open urban environments that attract creative workers. Today, tolerance must also encompass acceptance of non-human creativity. The study's data shows that frequent AI users exhibit greater comfort with machine-generated content, suggesting that exposure fosters cultural acclimation. However, this raises critical questions: Should the creative class tolerate AI as a peer? Or does tolerance now require resisting AI's homogenising effects?

Collectively, these implications necessitate an updated conceptualisation of Florida's framework. To properly address these paradigm shifts, this study suggests adding a fourth T, namely Technique, to the Creative Class theory, in order to capture the hybrid skill sets required for human-AI collaboration. Technique encompasses:

- AI literacy (understanding model limitations, bias mitigation).
- Adaptive creativity and innovation (blending traditional artistry with algorithmic iteration).
- Ethical and legal fluency (navigating copyright, attribution, and data sourcing).
- Active engagement in the new algorithmic age

4.2.3 Broader Implications

The widespread adoption of AI in creative work, coupled with stable, reliable high-bandwidth networks has reshaped the infrastructure needs of the creative economy. Traditionally, and previously necessitated by Florida's Creative Class theory, it was said that cities needed centralised creative hubs and fast public transport in order to facilitate collaboration and knowledge exchange. However, in the wake of this new technological era and COVID-19's impact, seamless remote collaboration is now widely adopted. For example, musicians no longer have to be in the same physical studio; instead, they can co-produce albums from different continents, diminishing the need for physical infrastructure. This shift demands a reimagining of urban planning and policy priorities. Instead of funding fixed creative hubs, policymakers might invest in flexible, temporal social spaces, such as co-working cafés, pop-up maker studios, or hybrid event venues, that cater to the human need for interaction without requiring permanent physical footprints. The post-COVID reality has shown that while remote work is viable, social isolation remains a critical challenge. Creative workers still require opportunities to build trust, exchange ideas, and foster collaborations; these are all functions that purely digital spaces struggle to replicate. For example, software engineers may work remotely 80% of the time but benefit from mandatory in-person meetups (e.g., one day per week) to maintain team cohesion.

The concerns regarding AI should also be addressed. As AI facilitates efficiency and changes in the structure of industries, Bachelor degree holders are fearful of the prospect of job displacement in the future, possibly leading to them moving away from creative professions. In order to deal with the fear of being deprived of career opportunities, policymakers should invest more in education curricula, creating new courses that embed AI toolsets directly into creative

courses at differing levels of schooling. This implementation can currently be seen in select institutions, such as the University of York and Nottingham Trent University (Ntu.ac.uk, 2025) (York, n.d.). This will prepare students for this new landscape, effectively dealing with insecurities modern students and professionals face. In essence, this is a recommendation for educationalists in order to prioritise this kind of learning and incorporate AI into their curricula.

5. Conclusion

This paper set out to redefine what being a creative worker means in an era of AI, aiming to capture the nuanced perceptions of creative professions by gauging their thoughts regarding AI's impact on authorship, ownership and occupational identity. By investigating the emerging skillsets and shifting professional landscapes, this study has updated Richard Florida's Creative Class theory. Empirical evidence challenges experience-based hierarchies, revealing that frequency hands-on engagement is the strongest driver and predictor of a positive AI perception. Furthermore, multivariate analysis has confirmed that practical exposure to AI is much more significant compared to other factors, suggesting that the solution to smoothly integrating AI into the CCIs lies not in philosophical debates, but in mandated, hands-on experience.

This study makes a significant theoretical contribution by refining and expanding Florida's "3T's" framework. Technology, as we know it, is no longer a mere tool for enabling creativity, but an active co-creator and gatekeeper, reshaping the demand for traditional skill sets. Consequently, talent is redefined from innate artistic skills to AI agility, in other words the ability to orchestrate and collaborate with intelligent systems. Furthermore, Tolerance must evolve from a geographical concept to encompass a social acceptance of machine authorship. To fully capture this paradigm shift, this paper proposes a fourth "T": Technique. Technique encompasses the hybrid competencies of AI literacy, adaptive creativity, and ethical fluency required to thrive in this new creative economy.

Broader implications also point to a profound need in restructuring existing spatial and educational structures. There is an increasing need for high-bandwidth networks over fixed, in-person creative hubs; this suggests that urban policies must shift to invest towards flexible, temporal social spaces (e.g. pop-up studios, co-working cafés) in order to mitigate the isolation of remote work while fostering real human interaction, which is essential for innovation. Another concern yet to be addressed is the fear of job displacement revealed amongst Bachelor's holders; thus, a primary practical implication for industry leaders and policymakers is to prioritise accessible AI tool training for emerging workers, as well as to create frameworks that encourage experimentation over theoretical discussions. Educational institutions are also encouraged to proactively embed AI toolsets directly into creative curricula at all levels, equipping future creatives to critically and ethically harness AI, future-proofing their careers and the vitality of the CCIs.

The limitations of this study suggest that this paper's findings are interpreted with caution. The modest sample size of 52 and the underrepresentation of certain demographics (e.g. Doctorate holders and traditional artisans) necessitate further validation. Furthermore, the

cross-sectional design merely captures a single moment in the rapid evolution of AI technology, where perceptions are inherently fluid and subject to change. Future research should therefore employ longitudinal, mixed-methods approaches with larger, more diverse global cohorts to track attitudinal shifts and explore the qualitative nuances of occupational identity negotiation. Despite these limitations, what this paper provides is an empirical foundation which asserts that the future creative class will be defined not by resistance, but by collaboration between human intuition and algorithmic power.

Works Cited

- De Cremer, D., Bianzino, N.M. and Falk, B. (2023). How Generative AI Could Disrupt Creative Work. [online] Harvard Business Review. Available at: <https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work>.
- Dembicki, G. (no date) Creative but Poor? Yep, the ‘Creative Class’ Has Its Class Divide | The Tyee. Available at: <https://thetyee.ca/News/2014/10/24/Creative-Class-Divided/#> (Accessed: 16 January 2025).
- Esling, P. and Devis, N. (2020) ‘Creativity in the era of artificial intelligence’. arXiv. Available at: <https://doi.org/10.48550/arXiv.2008.05959>.
- Florida, R. (2003) ‘Cities and the Creative Class’, *City & Community*, 2(1), pp. 3–19. Available at: <https://doi.org/10.1111/1540-6040.00034>.
- Florida, R. (2012). *The Rise of the Creative Class--Revisited*. Basic Books.
- Gao, M. and Leong, W. (2024). Research on the Application of AIGC in the Film Industry. *JOURNAL OF INNOVATION AND TECHNOLOGY*, [online] 2024(22), pp.2805–5179. Available at: http://eprints.intimal.edu.my/2027/3/joit2024_22b.pdf.
- Glaeser, E. (n.d.). Review of Richard Florida’s *The Rise of the Creative Class*. [online] Available at: https://scholar.harvard.edu/files/glaeser/files/book_review_of_richard_floridas_the_rise_of_the_creative_class.pdf.
- Hearn, G. (no date) ‘1. The future of creative work: creativity and digital disruption’.
- Lee, H.-K. (2022) ‘Rethinking creativity: creative industries, AI and everyday creativity’, *Media, Culture & Society*, 44(3), pp. 601–612. Available at: <https://doi.org/10.1177/01634437221077009>.
- Mahmud, B.U., Hong, G.Y. and Fong, B. (2022). A Study of Human-AI Symbiosis for Creative Work: Recent Developments and Future Directions in Deep Learning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(2). doi:<https://doi.org/10.1145/3542698>.
- Mántaras, R.L.D. (2017). *Artificial Intelligence and the Arts: Toward Computational Creativity*. [online] Csic.es. Available at: https://digital.csic.es/bitstream/10261/156712/1/FBBVA_TNS%282016%29_100-25.pdf.
- Nathan, M. (2008) ‘Creative Class Theory and Economic Performance in UK Cities’; in T. Yigitcanlar, K. Velibeyoglu, and S. Baum (eds) *Creative Urban Regions*. IGI Global, pp. 80–93. Available at: <https://doi.org/10.4018/978-1-59904-838-3.ch005>.
- Ntu.ac.uk (no date) *Mastering AI for Creative Practices*. Available at: <https://www.ntu.ac.uk/course/short-courses/art-and-design/mastering-ai-for-creative-practices> (Accessed: 31 August 2025).
- OECD (2021) *Economic and social impact of cultural and creative sectors: Note for Italy G20 Presidency Culture Working Group*. OECD Local Economic and Employment Development (LEED) Working Papers. Available at: <https://doi.org/10.1787/4d4e760f-en>.
- Saini, N. (2023). *Research Paper on Artificial Intelligence & Its Application*. *IJRTI*, 8(4).

- Small, Z. (2023) 'Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History', The New York Times, 4 July. Available at: <https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html> (Accessed: 12 December 2024).
- UNESCO (2021) Cities, culture, creativity: leveraging culture and creativity for sustainable urban development and inclusive growth. UNESCO, World Bank. Available at: <https://doi.org/10.58337/IIRA3223>.
- Wilson-Barnao, C. and Collie, N. (2024) 'Art rocks: Atmospheres of connection and everyday creativity in hybrid public space', *International Journal of Cultural Studies*, 27(3), pp. 405–427. Available at: <https://doi.org/10.1177/13678779231219016>.
- York, U. of (no date) MSc Artificial Intelligence for the Creative Industries, University of York. Available at: <https://www.york.ac.uk/study/postgraduate-taught/courses-2025/msc-ai-for-the-creative-industries/> (Accessed: 31 August 2025).

Designing Inclusive Digital Platforms for the Silver Generation By Sehyeon Kim

Abstract

The study examines the digital accessibility challenges that older adults face in South Korea using Baemin, the leading food delivery application. Given the rise in the aging population, understanding digital inclusion barriers helps promote digital equity. The study uses the human-centered design methodology to carry out the analysis and semi-structured interviews to identify barriers. Three primary themes emerged: unfamiliarity with digital interface, complex design formats, and the need for intuitive information delivery. The findings reveal that older adults are willing to utilize food delivery apps, but the barriers make it hard for them. The study recommends AI-assisted voice interfaces, simple visual hierarchies, and large design elements to bridge the gap.

Introduction

Problem Context

The global aging population is rapidly rising, and research projects a 2.1 billion increase by 2050. The aging population includes older adults above the age of 60 years. Some of the factors leading to the rise in the aging population include rising life expectancy and declining fertility rates. The other trend is the increased digitization of essential services, and the changes present an opportunity to support the aging population in reducing the exacerbating inequalities in access and use of digital tools. While the digital tools can offer potential benefits such as the convenience and access to essential services, addressing the accessibility, infrastructure, and training is key to ensuring they become a supportive tool for the older adults instead of an exclusion that affects their access (Mubarak & Suomi, 2022). My own grandmother is a clear example. Every time she wants to order food, she calls her children to do it for her. Even late at night, when she wants something to eat, she does not use the app; she walks outside to buy it herself. I have seen and heard this many times, and it made me realize something important: technology is not truly inclusive if people like my grandmother cannot use it. The inspiration informs a deliberation of how delivery apps could work better for the elderly. Designing a more user-friendly version of a food delivery app using AI and voice recognition can help older adults access essential services.

Research Focus and Significance

The paper focuses on food delivery in Korea, which has become the cornerstone of lifestyle for most people's everyday lives. Baemin has emerged as an industry leader as it shows how digital platforms transform convenience. It also highlights the broad challenges of digital inclusion when it comes to catering to the needs of all populations, especially the vulnerable, such as older adults. Younger users have an easy time using the application, while the older clients face barriers that limit digital economy participation. One of the barriers is the interface

complexity, as elders struggle with the small fonts due to physical changes such as decreased vision. Having small fonts, unfamiliar navigation structures, and dense menus affects the older adults' use of the application. Limited digital literacy affects knowledge of basic functions, including registering payment methods or applying for discounts. According to Mubarak & Suomi (2022), other issues include trust and security, as older adults may fear making mistakes or being fraud victims. Access to the digital essential services is thus determined by digital literacy and not physical proximity. The paper explores how using the human-centered design principles is important in ensuring greater accessibility to the FDAs. It helps to develop a simplified visual user interface and onboarding processes that empower users to utilize the application better. Designing with empathy also helps ensure that the solutions implemented address the lived realities of the users (Jiang et al., 2024). Taking advantage of emerging technologies, such as artificial intelligence and voice interaction, helps bypass traditional text navigation, making the experience smooth (Mubarak & Suomi, 2022; Rodriguez-Martinez & Thompson, 2024). The exploration is key to illustrating how digital transformation intersects with social inclusion. It lays the basis for the development of other applications so that they can be inclusive, including ease of use among older adults.

Background and Literature Review

Demographic Changes

The world is changing with the rapid development of medical and industrial technology. Human life expectancy has increased significantly, and with constantly falling birth rates, the global population has now been skewed towards the elderly. Research indicates that the 21st century will be shaped by elders, and some predict that in 2050, the population of the older generation will reach about 2.1 billion. “Silver generation” is a term that describes the elderly, those aged 65 and above. In Korea, the shift also occurs as the aging population increases. Research shows that there has been a 20% increase in the population from 2023, which has 9.73 million older adults (Yoon, 2025). The population represents 19% of the general population, showing its significance in designing products. South Korea is also expected to become the world's most aged population by 2045, surpassing Japan (Yoon, 2025). The silver generation's increase has caused demographic changes and new waves across industries, including the food sector. Older adults face challenges given the physical and cognitive decline changes. The declining vision, motor skills, and unfamiliarity with new technologies fully affect their ability to participate in the digital economy. Special consideration for such a population is thus key when developing digital tools to ensure inclusion.

The growth of Food Delivery Apps (FDAs)

Food Delivery Apps (FDAs) are becoming increasingly popular today due to the convenience they offer to the population in accessing quality food. These are platforms that clients can access through mobile applications or websites. The individual can order food from a

wide selection of food options and providers from the food vendors and local restaurants. The applications display menus, prices, and reviews to allow customers to choose food and pay for their order, and then they can receive their food within a stipulated time (Hoang & Le Tan, 2023). The industry has grown significantly in developed and developing nations, and Korea is among the nations that have adopted the technology. In Korea, an App called Baemin(배달의민족) is one of the most widely used food delivery apps. It allows users to choose a restaurant, select menu items, enter their address, and complete payment with just four steps. The system is fast, efficient, and deeply targeted towards modern urban society.

Baemin (배달의민족) is known for its sleek interface and the vast array of restaurants that people can choose from when ordering. It contains over 140,000 eateries and offers everything from sushi to fried chicken to vegan options, thus becoming the most preferred platform. Its scope covers more than the Korean market, allowing foreigners to log in using their overseas numbers and access the services. It accepts international payment methods, making it accessible to all populations. Users can also track their deliveries in real time, ensuring the food reaches them on time. Some of the attributes that best describe the app include its speed, reliability, and convenience.

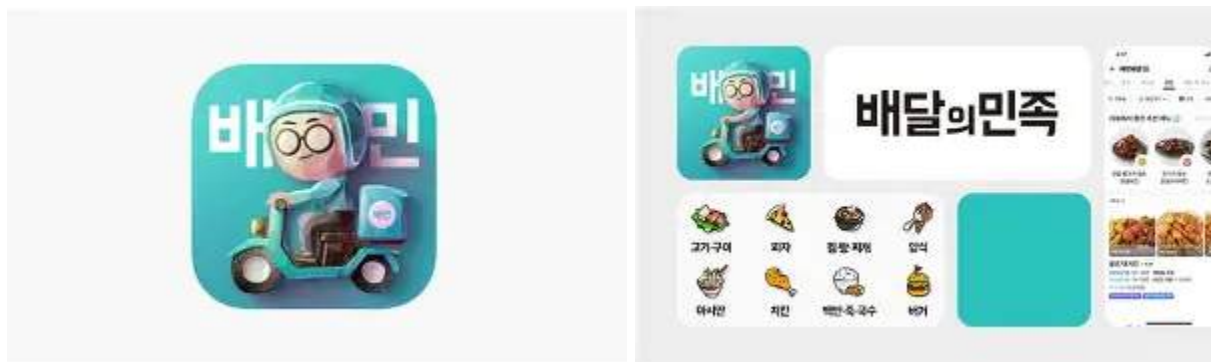


Figure 1: Baemin App (Li, 2025)

The Benefits and Constraints of FDAs

FDAs play an important role in addressing food insecurity issues among the population, especially older adults. A study by Hwang et al. (2024) recognizes the role that technology-based solutions, such as online food delivery (OFDs) and Food delivery Apps (FDAs), play in addressing issues such as food insecurity. According to the research, the applications since the pandemic have increased access to food, especially among older adults with constrained physical abilities, ensuring they receive proper nutrition. However, barriers exist to using food delivery apps that create inequalities in access to essential services for older adults. Chung-Heon Lee et al. (2024) studied the interaction and design barriers affecting older adults. The findings showed that a lack of a clear roadmap to guide the users through the purchase process, information overload, and poor interactive features are affecting the usage of the applications. Also, the majority of the elderly are unfamiliar with smartphones or touchscreen interfaces, which makes them feel overwhelmed and inaccessible. Due to failure in understanding app layouts (such as

searching, making a payment decision, selecting a menu, or resulting), the population faces multiple barriers.

Baemin may seem simple on the surface—typically involving four basic steps: selecting a restaurant, choosing menu items, entering an address, and completing payment. However, in those steps below, numerous sub-processes make the experience complex, especially for older users. For instance, after typing in the name of a restaurant, users are presented with a long list of similar options. Then they should browse through dense menus, small images, and tight texts. After selecting the restaurant and menu, they have to add them to a cart. Current formatting features small fonts and cluttered layouts, which make it difficult for elderly users to read and navigate accurately and easily, as shown in Figure 1. In addition, the app uses terminology that may be unfamiliar to older adults. Words like “최소주문” (minimum order amount) or “픽업가능” (available for pickup) are common in delivery app interfaces but are not part of everyday language for many seniors. These unfamiliar terms contribute to confusion and act as another barrier to access. This gap highlights a growing issue of isolation of the silver generation in a digitalized society, where digital solutions dominate the market.

Inclusive Design and Accessibility Principles

Various principles govern the universal design that can help in enhancing app usability among older adults. The universal principles include making the application simple and intuitive, providing perceptible information, and flexible use (Jiang et al., 2024). The app needs to be easy to understand and use, using a simple view. When it comes to perceptible information, the principle calls for having essential information, such as the order and menus, communicated effectively, regardless of the person's sensory abilities. Flexibility looks at the application's ability to accommodate various user preferences, such as adjusting the text size or fonts. It is also vital to consider the age-friendly design best practices that consider various factors, including navigation, text and color, error prevention, and imagery (Jiang et al., 2024). The app needs clear navigation, such as larger buttons for older adults, and to reduce clutter. Having a high color contrast for text can help create readable fonts, improving readability. Regarding error prevention, the confirmation prompts play an important role in allowing users to correct mistakes and reduce frustration easily (Tsai et al., 2023).

It is important to incorporate inclusive design and accessibility principles to cater to the needs of older adults. Research by De Silva et al. (2024) shows that using emerging technologies, such as Food-Talk, a groundbreaking voice-based online food delivery system, helps to meet the diverse needs of the population, including the visually impaired. The tool uses AI technology to allow users to use voice in registration and menu management, making the use of applications easy for older adults. Some of the benefits of using such technologies include increased accessibility, usability, and performance, which promote client satisfaction (Sayago & Blat, 2024).

Methodology

The human-centered design played an important role in exploring the issue and determining some of the challenges that older adults experience when using the Baemin application. A three-phase human-centered design methodology was used, including empathy and discovery through user research, problem identification and analysis, and ultimately ideation and concept development.

Current Analysis of the Application

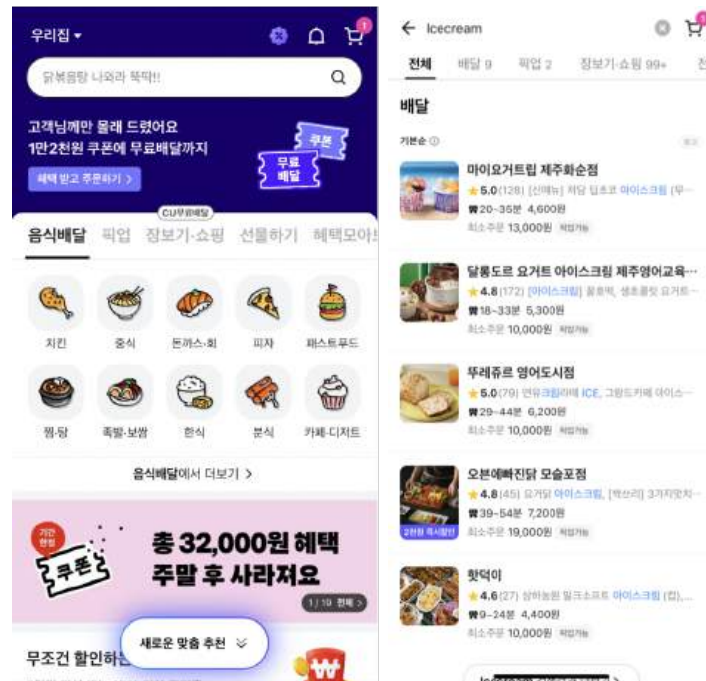


Figure 2: screenshot of Baemin App, showing the design of the app

The current application analysis phase revealed various gaps within the application. It involved analyzing the application to identify potential usability barriers affecting older adults. Various potential issues were identified, including the dense layout of the application, the use of small fonts, the complex navigation structures, and the terminology that is unfamiliar to the silver generation. For instance, terms like "최소주문" (minimum order amount) and "픽업가능" (available for pickup) point to common delivery app language, but may not be part of the everyday silver generation terminology. The app also seems to prioritize information density and efficiency, which overshadows clarity and simplicity, and assumes that all users are young and tech-savvy. There are currently four primary steps included in the ordering process, including restaurant selection, menu browsing, address entry, and payment process. There are, however, numerous sub-processes within the larger ones that affect its usage, making it strenuous.

User Interview Process

Semi-structured interviews were conducted with the silver generation population to gather insights into their experience with the application. The primary focus of the inquiry was to gain an understanding of the current food ordering behavior and technology comfort levels. It also included identifying the challenges they encountered while using the application. The recruitment took place at the community centers and through personal networks. The strategy ensured that the research obtained data from a diverse population that uses digital tools to order food. The interview questions targeted different areas, including traditional food ordering methods, preferences, attempts to use delivery apps, difficulties encountered, emotional responses to tech tools, and interaction methods. The open-ended questions allow users to provide detailed narratives that help collect rich data.

Thematic Analysis

Thematic analysis helped identify the patterns and core issues affecting the usage of food delivery applications. The process entailed multiple readings of the interview transcripts, identifying the significant statements, and grouping the experiences together into broad themes for understanding. The inductive approach was key to identifying themes from the data rather than imposing predetermined categories. The research process was tailored to the Korean population, given that it prioritized the use of participants' own language and explanations, ensuring that the themes reflected users' lived experiences. Moreover, the analysis followed the design thinking methodologies, ensuring it was grounded in user-identified needs and preferences.

Findings

The data collected during the interview process identified three major themes. They included unfamiliarity with digital interfaces, complex design formats, and the need for intuitive information delivery.

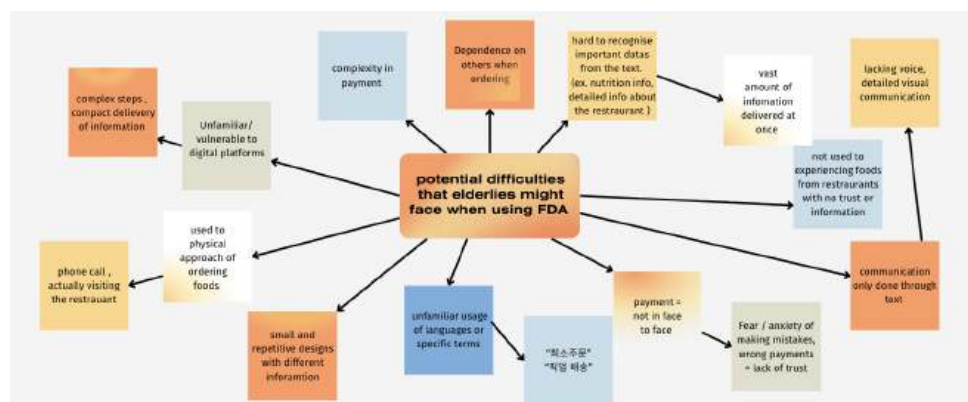


Figure 3: visual mindmap showing summary of potential difficulties that elderly people might face when using FDAs.

Theme 1. Unfamiliarity with digital interfaces

The results revealed that the older adults consistently expressed discomfort and unfamiliarity when it comes to the use of smartphone-based food ordering applications. Instead, they preferred traditional methods that they had used for decades. One of the participants said, “I usually have the restaurant call me directly, but not with these apps”. There was an insistence on verbal communication, which reflected deeper comfort with familiar interaction patterns. The unfamiliarity went beyond simple technology to other fundamental differences in information processing preferences. Users showed frustration when it came to navigating complex visual interfaces without guidance. Many respondents reported relying on family members to place orders, highlighting the role that social support plays in using the digital tools. The individuals who attempted to use delivery apps reported anxiety about making mistakes. One noted, “I am scared that if I press the wrong button, the money might go to the wrong place”. Such fears, which were expressed, create psychological barriers that prevent us from becoming older adults.

Theme 2: Complex Design Formats

The current Baemin interface presented numerous design challenges for the users. One of the identified issues was the small fonts, which made the interface dense and confusing for adults with visual issues. There are also cluttered layouts and information density, which prioritize efficiency but at the expense of clarity. The need for the app to present maximum information on a single screen created cognitive overload for the older adults, who needed more time to process the information. Restaurant selection, particularly, was a bit challenging. One of the respondents said, “I found it hard to tell which location has good food or is known for certain dishes as I was presented with a long list having similar options”. Also, in presenting the restaurant information, smaller images were used with abbreviated descriptions, making it hard to develop clear distinctions among the elderly users. Menu navigation was also an issue, given that participants showed difficulty distinguishing between the different items and understanding the pricing. The use of a tight visual format only appeals to the younger users; therefore, for the older adults, it creates accessibility barriers.

Theme 3. Need For Intuitive Information Delivery

The third theme focused on the intuitive information delivery within the app. Most participants expressed a clear preference for information that is presented with methods aligned with their existing mental modes and experiences. They mentioned terms like “large”, “clear images”, reminiscent of the traditional paper menus, which were preferred over the small thumbnails. Their outcomes also revealed that the elderly prefer processing information more effectively when it follows familiar patterns and offers clear guidance. Therefore, breaking down complex processes into simple sequential steps with obvious progress indicators will help them use their applications effectively. It also reduces cognitive overload and increases confidence when it comes to app usage. Having verbal instruction and feedback was also important, as it arose as a key theme. Various participants expressed comfort with spoken information that could be processed at their own pace, rather than dense information. The finding was therefore

important in pointing to the fact that voice-based interfaces have the potential to improve communication among older users.

Discussion

The research findings showed a significant gap between the current food delivery application design and the needs of older adults. It also suggested opportunities for more inclusive design approaches that could help expand these applications' markets and address the digital equity concern. One identified barrier is the visual interface that affects older adults due to their cognitive overload and difficulty processing the information. Studies show that an enhanced AI-assisted voice interaction can help address this barrier, allowing the older adults to interact effectively with the application (De Silva et al., 2024; Sayago & Blat, 2024). It also mirrors the traditional phone ordering experiences, making it familiar to the users. It is also important to address the visual issue by having an enhanced visual Information architecture, which helps provide larger images and straightforward typography that offer a simple layout for the users (Juliá-Nehme & Rosell, 2024). The use of AI design incorporates clear communication patterns and targets older adults, ensuring that the application is suited for their needs and preferences. The solution helps to maintain agency by offering multiple interaction modalities rather than completely replacing the traditional interface (Liusayago et al., 2023). The users can therefore choose whether to use the voice commands, the touch interface, or a combination for their maximum use, creating flexibility.

The research contributes to inclusive design literature as it demonstrates the value of human-centered research methods in understanding how all adults adapt and use technology (Phil Klaus et al., 2024). The findings also challenged the common assumptions about the elderly's technology aversion, revealing that with some adjustments, the older adults are willing to adopt digital technologies. If designers design interfaces that have elderly users' needs in mind, such as cognitive processing, adults are more likely to use digital tools (Hwang et al., 2024). The other important realization is the potential of voice interface technologies in bridging the digital divide among the populations (Sayago & Blat, 2024). Rather than requiring older adults to adapt completely to a new interaction paradigm, using the AI assistant approach helps them use their existing communication skills and introduce digital convenience gradually to users.

Practical Application of the Findings

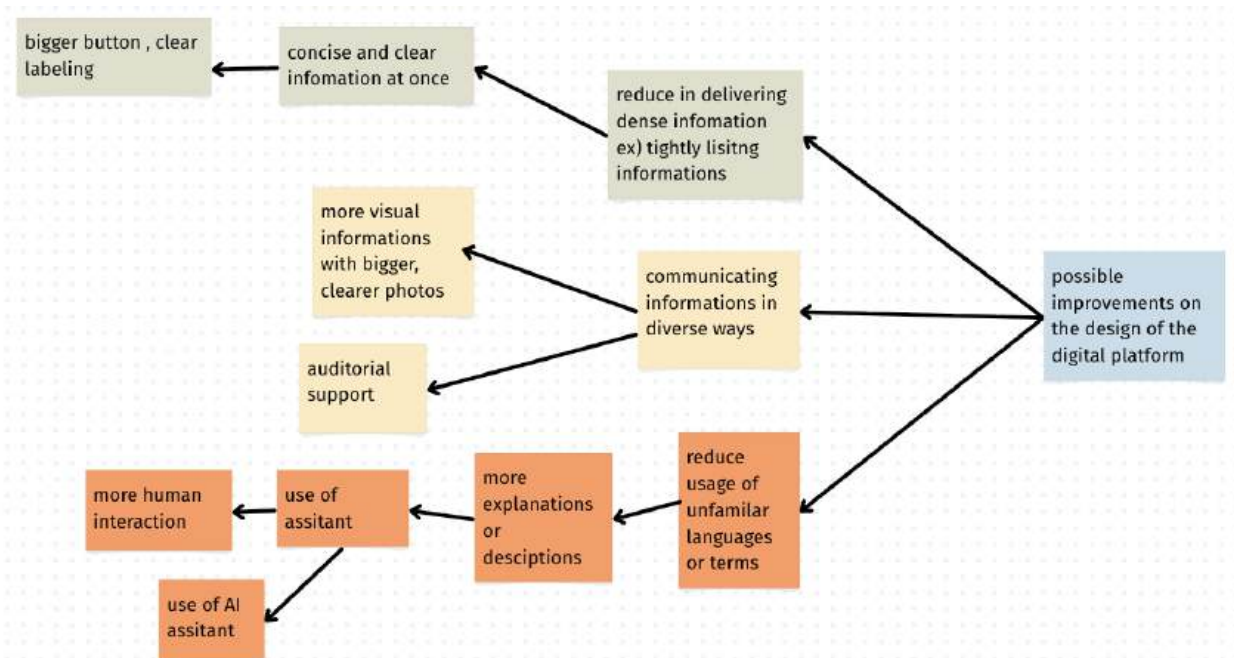
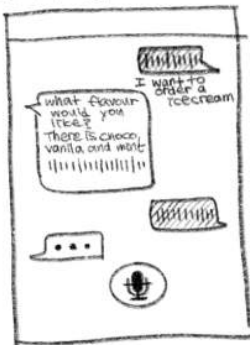


Figure 4: visual mindmap showing possible improvements on design of digital platform in order to support elderly users.

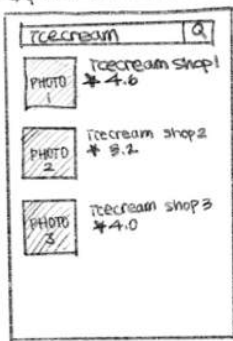
The design framework developed through the research offers practical guidance for food delivery companies and other digital service providers. This can help them expand accessibility among the different populations, therefore expanding their market reach. Adopting the inclusive design extends beyond the social responsibility to market expansion opportunities. Some of the implementation considerations include incorporating the technical development of voice recognition systems that are optimized for older adults' speech patterns, integrating an AI assistant with existing app architectures, and training their staff for customer service representatives who may need support for elderly users.

Final proposal of Design

① Through voice memo communication with AI

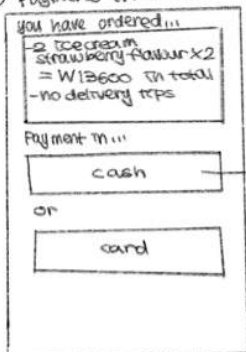


② expand the visual information



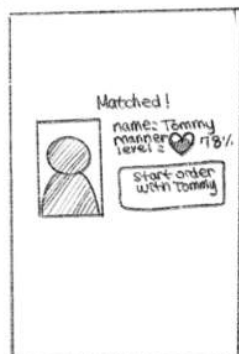
by clicking this button it shows different shops options

③ Payment in cash.



payment system more familiar to elders → face to face payment

④ Matching a helper / assistant



through a short call human assistant supports elders to order the food

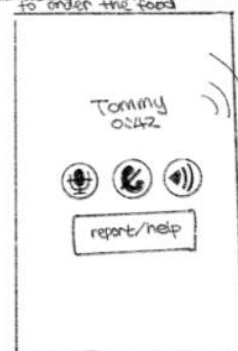


Figure 5 : initial ideation of the design showing sketches of storyboards



.Figure 7: visualising the final design using app “Figma”

Limitations

The study had various limitations that may affect its applicability. One was a small interview sample size of 20 participants that focused on a single cultural context within Korea and targeted one app, Baemin. The results are therefore relevant for the Korean population and food industry. Also, while the study findings show valuable insights into the use of delivery apps by older adults and their experiences, there is a need for broader validation through large-scale studies and cross-cultural research. Future research should therefore include prototype testing with elderly users to validate their proposed design solutions and gather feedback.

Conclusion And Recommendation.

One of the findings is that the difficulties of elderly users stem from the focus of the applications on accessibility efficiency. Using the human-centered design approach reveals specific barriers such as unfamiliarity with digital interfaces, complex visual formats, and the need to have intuitive information delivery systems. Therefore, key recommendations for industry practitioners are to ensure that they invest in accessibility research and inclusive design practices that consider the older adult users from the initial design phase. When it comes to the design features, it's important to incorporate larger visual elements and simplified information hierarchies. This helps to ensure that the older adults are familiar with data technology and have multiple interaction modalities, including voice interfaces, visual, or both. When implementing the strategies, the companies should also adopt gradual rollout approaches that allow the older adults to transition comfortably to the digital platform and also have an alternative ordering method during the adoption period. Customer support systems are also important in enhancing and providing step-by-step guidance to the users.

The research contributes to growing recognition that inclusive design benefits all users and addresses the digital equity concerns. Given the rise in digital platforms in today's era, there is a need to ensure accessibility for older adult users as a business opportunity and a social responsibility. The study has demonstrated that combining interface analysis, user interviews, and iterative design development helps develop. A replicable framework for addressing digital inclusion challenges across the different service sectors. Also, integrating traditional usability principles with emerging voice interface technology will help to offer a promising direction for the future.

Works Cited

- Chung-Heon Lee, Yuna Hwang & Hayeon Song. (2024). Designing Conversational Agents for Older Adults: Effects of Conversational Form and Nonverbal Factors on Mobile Banking Experiences. *International Journal of Human–Computer Interaction* 0:0, pages 1-14.
- De Silva, D.I., Godapitiya, M.V.N., Peiris, M.M.A.E., Perakum, K.K.P., Mangchanayaka, M.V.V., Senevirathne, U.W.H.N. (2024). Food-Talks: Food Ordering and Delivering Application for Visually Challenged. In: Yang, X.S., Sherratt, S., Dey, N., Joshi, A. (eds) *Proceedings of Ninth International Congress on Information and Communication Technology. ICICT 2024* 2024. *Lecture Notes in Networks and Systems*, vol 1013. Springer, Singapore. https://doi.org/10.1007/978-981-97-3559-4_16
- Hoang, H., & Le Tan, T. (2023). Unveiling digital transformation: Investigating technology adoption in Vietnam's food delivery industry for enhanced customer experience. *Heliyon*, 9(9), e19719. <https://doi.org/10.1016/j.heliyon.2023.e19719>
- Hwang, S., Johnson, C. M., & Charles, J. (2024). Food Delivery Apps and Their Potential to Address Food Insecurity in Older Adults: A Review. *International Journal of Environmental Research and Public Health*, 21(9), 1197. <https://doi.org/10.3390/ijerph21091197>
- Jiang, Q., Deng, L., Zhang, J., & Pengbo, Y. (2024). User-Centered Design Strategies for Age-Friendly Mobile News Apps. *SAGE Open*. <https://doi.org/10.1177/21582440241285393>
- Juliá-Nehme, B., & Rosell, J. (2024). Interaction and Design Barriers for Older Adults in Food Delivery Apps: A Usability Study. *International Journal of Human–Computer Interaction*, 41(9), 5761–5778. <https://doi.org/10.1080/10447318.2024.2365484>
- Liu, M., Wang, C., & Hu, J. (2023). Older adults' intention to use voice assistants: Usability and emotional needs. *Heliyon*, 9(11), e21932. <https://doi.org/10.1016/j.heliyon.2023.e21932>
- Mubarak, F., & Suomi, R. (2022). Elderly Forgotten? Digital Exclusion in the Information Age and the Rising Grey Digital Divide. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*. https://doi.org/10.1177_00469580221096272
- Phil Klaus, Karen Edwards, Daniela Norvik, Aikaterini Manthiou, Van Ha Luong; Inclusive servicescapes: the imperative of universal design principles. *Journal of Services Marketing* 19 November 2024; 38 (8): 1089–1098. <https://doi.org/10.1108/JSM-04-2024-0150>
- Rodriguez-Martinez, A., & Thompson, K. (2024). User-centered AI-based voice-assistants for safe mobility of older people in an urban context. *AI & Society*, 39(2), 445-462. <https://doi.org/10.1007/s00146-024-01865-8>
- Sayago, S., & Blat, J. (2024). Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction*, 27(4), 1-27. <https://doi.org/10.1145/3373759>

- Tsai, T. H., Tseng, K. C., & Chang, Y. S. (2023). Design guidelines of mobile apps for older adults: Systematic review and thematic analysis. *JMIR mHealth and uHealth*, 11, e43186. <https://doi.org/10.2196/43186>
- Yoon, L. (2025). Aging population in South Korea - Statistics & Facts - Statista, Demographics. Available at: https://www.statista.com/topics/12258/aging-population-in-south-korea/?__sso_cookie_checker=failed (Accessed: 05 September 2025).

Evaluating Psychedelics in Depression Treatment: A Comparative Review of Ketamine, Mescaline, and Psilocybin By Hanayo Endo

Abstract

Depression is a leading mental health concern that affects millions globally and is commonly treated using psychotherapy, antidepressants, or both. However, psychotherapy is costly and antidepressants cause adverse effects such as gastrointestinal issues, sleep disturbances and weight changes. In recent years, psychedelics have regained popularity as a focus of research after being illegal for years, as they offer a faster, more effective therapeutic effect than antidepressants, especially in patients with treatment-resistant depression. Psychedelics, or hallucinogens, are psychoactive substances that alter sensory experience, perception, thought processes, mood and consciousness by binding to receptors in the brain. These effects have been vital in religion for healing rituals, and have gained the attention of medical researchers as a treatment option for depression, with much research showing promising results. Clinical and naturalistic research on psychedelics such as mescaline, ketamine, and psilocybin were found to improve depressive symptoms quickly, without having long-lasting adverse effects. This review examines and evaluates the existing evidence for three psychedelics (ketamine, mescaline, and psilocybin) and compares the effectiveness and tolerability as an emerging depression treatment. Ultimately, this comparative review finds that all three of the psychedelics show effectiveness in alleviating depressive symptoms, psilocybin is considered the most effective psychedelic in treating depression due to its rapid, long-lasting improvement and safety. By examining research on psychedelics, understanding the comparative benefits and identifying areas that require more research, it allows for improvement in the future of psychedelic-assisted mental health treatment and may help individuals who are unresponsive to conventional treatment methods.

Keywords psychedelics, depression, ketamine, mescaline, psilocybin

Introduction

Globally, an estimate of 332 million people suffer from depression (“GBD Results”). Depression is a multifactorial disorder that has a wide range in symptoms that are persistent and interfere with daily lives. These symptoms can be categorised into behavioural and motor symptoms (agitation, psychomotor retardation, suicide attempts, negative verbalisation), cognitive symptoms (negative self-view, pessimistic views, hopelessness about future, lack of motivation), social symptoms (withdrawal from social or recreational activities) and biological symptoms (change in weight, fatigue, loss of energy, reduced appetite). Unlike sadness, which is temporary, depression is a condition that is persistent within the emotional domain that disrupts daily life of individuals (Rondón Bernard).

To treat depression, antidepressants and psychotherapy are commonly utilised. Antidepressants are drugs that are commonly used in the field of mental health to help alleviate

the depressive symptoms. Most antidepressants primarily increase the levels of neurotransmitters associated with mood regulation by preventing reuptake, blocking inhibitory receptors, or disabling enzymatic breakdown (Artigas et al.). Psychotherapy or talk therapy is a classic method used to treat depression where patients talk to therapists to identify and change negative thought patterns (Rogers, Carl R.). However, both treatment methods have flaws. Patients taking antidepressants may have to control their diet and can experience adverse, sometimes severe, side effects (Ramic et al.; Braund et al.) as well as becoming overly reliant on the drug, as the brain gets adapted to producing unnatural levels of neurotransmitters such as serotonin and dopamine (Fava). Psychotherapy is less accessible and costly compared to drug treatments and requires high levels of motivation for effectiveness to be evident. Furthermore, it is possible that patients are resistant to both or either of these treatment methods and it is estimated that between 30 to 55% depressed patients have treatment-resistant depression (McIntyre et al.) which is defined as the failure to respond to two or more antidepressant trials (Gaynes et al.). To treat individuals with (treatment-resistant) depression, there has been an increase in research surrounding psychedelic-assisted therapy (Breeksema et al.). Many studies have demonstrated the potential benefits associated with psychedelic-assisted therapy for depression such as the rapid improvement and reduction of adverse depressive symptoms as well as the minimal, tolerable side effects. For example, a meta-analysis that investigated 8 research on psychedelic-assisted therapy for depression found depressive symptoms decreased greatly from the first day up to 6 months after psychedelic was taken (Ko et al.). Another meta-analysis analysing 14 studies found similar results (Romeo et al.), hence, these findings suggest that psychedelics have the potential and could be a possible alternative treatment method for depressive individuals especially in those with treatment-resistant depression (Kalfas et al.).

Psychedelics have been used by humans for thousands of years, with the first use of the drugs being in indigenous cultures found in America, Africa, and Asia. Psychedelic plants and fungi such as psilocybin mushroom, peyote, and ayahuasca, were used primarily for spiritual, healing, and rituals as a way to connect with the gods (supreme beings), treat illnesses, or ceremonial uses. The first modern psychedelic research was done by Albert Hofmann in 1938 when he synthesised LSD (lysergic acid diethylamide). In 1943, Hofmann became the first human to experience the psychedelic effect when he ingested LSD by accident. This was documented as his “bicycle ride” and marked the start of the second phase of psychedelic research and exploration. The golden research age of psychedelic started in the 1950s to the 1960s, where psychedelics were widely studied by psychiatrists, psychologists, and researchers as a treatment for mental health disorders such as alcoholism, depression, anxiety, and trauma. In this time, psychedelics gained the attention of the wider society and became central to the countercultural movement that emphasised spiritual exploration, creativity, opposition to normalised societal norms. Due to the countercultural movements, psychedelics became associated with anti-war protests, hippie culture, societal unrest, and was criminalised by governments worldwide. As a result, psilocybin, LSD, mescaline, and other psychedelic drugs were classified to have “high abuse potential” and were banned in the medical field by the

Controlled Substances Act in the 1970s U.S. However, after the 1990s, advocacy and careful research led to the doors of psychedelic studies to reopen and since then, rigorous clinical trials have been carried out. Moreover, the FDA Breakthrough Therapy Designation was recently granted to psilocybin and MDMA, which marked a significant turning point towards the potential legalisation of psychedelics use in the medical field (Belouin and Henningfield; Jay).

Psychedelics are often referred to as hallucinogens that alter behaviour in humans (Kyzar et al.). Classic serotonin psychedelics include psilocybin, mescaline, lysergic acid diethylamide (LSD) and N,N-Dimethyltryptamine (DMT), and are serotonin receptor (5-HT_{2A}) agonists. An agonist is a drug that targets a specific receptor by binding to the receptor site and activating the receptor to induce a biological effect (amplifies the effect of the neurotransmitter) (Muir and Sams). The mechanism of psychedelics is extremely complex, however, it is said that classic serotonin psychedelics affect the serotonin system by binding to the brain's serotonin 5-HT_{2A} receptor and stimulating them. There is also evidence that suggests classic serotonin psychedelics increase the neuroplasticity, formation of neural connections (Husain et al.). The activation of serotonin receptors promotes dendrite spine and synapse growth, increasing communication between different brain regions.

Ketamine is widely used as a dissociative, short-acting anaesthetic, relieving pain and inducing unconsciousness before and during surgery in animals and humans. However, it is considered a psychedelic due to the hallucinogenic effects or psychedelic experiences it induces. Unlike classic serotonin psychedelics, ketamine is an antagonist, a chemical that binds to and blocks receptor sites of a neurotransmitter, and affects the glutamate system in the brain. As a result, antidepressant effects are induced when there is an increase in concentration or surge of glutamate, an excitatory neurotransmitter, in the brain (Savalia et al.). The glutamate surges promote the formation of new synapses or neuroplasticity, connecting neurons that do not usually connect. This allows thinking to become more flexible, negative thought patterns to be broken and strengthen positive neural connections (Kim et al.).

Through the exploration of a number of studies, this review examines and evaluates the existing evidence from both clinical trials and natural setting of depression treatment, utilising three specific psychedelics: ketamine, mescaline and psilocybin. It further compares the effectiveness and tolerability of the three psychedelics as an emerging depression treatment. It concludes that psychedelic-assisted therapy is effective in treating various types of depression such as Treatment-Resistant Depression and Major Depressive Disorder, with mescaline showing great promise in naturalistic, unsupervised settings, ketamine offering rapid relief, and psilocybin providing long-lasting, sustainable benefits in adults.

By assessing the existing research on psychedelics and highlighting the comparative benefits, areas that require more research can be identified. This review intends on contributing to the advancement of psychedelic-assisted mental health treatment and allows for the exploration of potential options to help individuals who are unresponsive to conventional treatment methods.

Mescaline

Mescaline is naturally produced psychedelic by the Mexican peyote cactus and is found in the phenethylamine class (Stork and Schreffler) that was explored and suggested to have potential therapeutic effect and application in 1936 (Guttman). Mescaline is a partial agonist at serotonin receptors and binds to 5-HT₂ serotonin receptors, producing hallucinogenic effects when activated in the prefrontal cortex. This hallucinogenic effect is caused by the release of glutamate and excitation of neurons which amplifies the signals that are being sent to other brain regions, altering the cortical and limbic network, heightening sensory and emotional processing (Vamvakopoulou et al.; Davis).

Additionally, the activation of 5-HT₂ serotonin receptors promotes neuroplasticity by promoting the growth of new dendrite spines and synapses as well as increasing communication between different brain regions. In depression, certain brain circuits become constrained and overconnected. Therefore, psychedelics like mescaline allow these circuits to reset to become more flexible in thinking and emotional processing, disrupting the rigid, negative thought patterns and offering individuals with new perspectives.

A study done by Agin-Liebes et al. utilised an anonymous online questionnaire, aiming to collect self-report data on the use of mescaline in a naturalistic, non-clinical setting amongst 184 depressive adults. Amongst the 184 depressive adults, 86% reported improvements in their depression after using mescaline. Those who reported an improvement in depression would often give relatively high ratings of acute psychological factors such as mystical-type, psychological insight, and ego dissolution after using mescaline. These acute psychological experiences when using mescaline were associated with the increased number in self-reported improvement in a number of domains of depression such as mood and anxiety. Therefore, it was concluded that mescaline is an effective treatment option for depression by improving depressive adults' mood and providing therapeutic benefit in natural, unsupervised settings.

Similarly, Uthaug et al.'s study also aimed to investigate the naturalistic use of mescaline, using an online survey to collect data on the patterns, motivation, acute effects and long-term outcomes when using mescaline in 452 respondents, including some reporting depression. Through the survey, it was found that most of those with depression reported improvement after using mescaline, with some acute mystical type effects that were tolerable. Hence, Uthaug et al. concluded that mescaline can significantly improve depression in a natural, unsupervised environment. Moreover, mescaline has minimal adverse effects and low potential for abuse, thus mescaline is an effective and safe treatment option for depression, thus supporting Agin-Liebes et al.'s findings.

While the two previous studies utilised online methods to gather large amounts of data, Halpern et al. used method triangulation (screening interview, substance-use history, psychiatric history and neuropsychological tests) to investigate the after effect of long-term mescaline use by comparing three groups of people from the Navajo Native Americans: 61 Native American Church members who are regular peyote (mescaline) users, 36 with a history of alcohol dependence but has been sober for 2 or more months, 79 individuals who reported minimal use

of substances including peyote and alcohol. Results found that long-term peyote users had better depression and mental health scores than the control group. The findings of Halpern et al. suggest that long-term peyote (mescaline) use was not linked to higher levels of depression. Instead, the longer participants used peyote, the lower the self-reported depression scores were. This indicates that structured, religious mescaline use can improve psychological well-being, such as reduced depressive symptoms.

Collectively, these studies on mescaline-assisted depression treatment demonstrate that mescaline is functional in treating depression, with many users reporting significant improvement after use. However the study does not provide evidence that mescaline is an effective clinical treatment for depression. Instead, it highlights that mescaline is a safe treatment depressed individuals may use in natural and unsupervised settings due to its low potential for abuse and minimal adverse effects. Additionally, these studies (Agin-Liebess et al.; Uthaug et al.; Halpern et al.) emphasise the need for further clinical trials to determine whether the relationship found reflects the pharmacological effects of mescaline or other factors related to its use such as the context of mescaline/peyote consumption.

While mescaline's positive effects are seen in naturalistic and unsupervised settings, the next section focuses on ketamine, unlike classic psychedelics, is an antagonist with rapid, clinically tested antidepressant effects that show a slightly different potential in psychedelic-assisted therapy for depression.

Ketamine

Ketamine has been gaining attention as a potential treatment for depression and its hallucinogenic properties such as psychedelic experiences and hallucination when used, classifies them as psychedelic. Ketamine is an antagonist, a chemical that binds to and blocks receptor sites of a neurotransmitter, and affects the glutamate system in the brain. Glutamate is an excitatory neurotransmitter, therefore it increases the firing of postsynaptic neurons after binding to its complementary receptor sites (Zanos and Gould). However, ketamine blocks this glutamate receptor, the N-methyl-D-aspartate (NMDA) receptor, which increases the firing of a type of excitatory nerve cell in the brain known as the glutamatergic pyramidal neuron to cause glutamate surges. It is believed that the increased concentration of glutamate leads to the antidepressant effects of ketamine as neuroplasticity is promoted, creating new neural connections which allows more flexibility in emotional processing (Kadriu et al.; Artin et al.).

Although the glutamate surges induced by ketamine can have positive antidepressant effects, ketamine can also cause pleasurable, rewarding effects, hence, the drug has been used as a "club drug" for a long time. Ketamine blocks NMDA receptors in the brain, increasing dopamine levels in areas of the brain linked to reward (Liu et al.).

A five year longitudinal study done by Gutierrez et al. investigated the effectiveness of low-dose infusions of ketamine in treating Major Depressive Disorder with Treatment-Resistant Depression (MDD-TRD) in 71 outpatients. Results showed that low-dose infusions of ketamine had a 54.93% response rate with rapid and effective reduction in depressive symptoms and

suicidal ideation. Furthermore, 78.26% of patients experienced mild, temporary and tolerable side effects. These findings suggest that ketamine has potential in providing individuals with depression, especially MDD-TRD, a rapid, safe and effective treatment that significantly reduces depressive symptoms and suicidal ideation with minimal adverse effects.

Another similar study done by Mandal et al. also investigated the effectiveness of ketamine. However, instead of infusions, this study injected participants with six low-doses of ketamine over two weeks in 25 patients of the male sex with severe depression. An hour after receiving the first ketamine injection, there was a significant reduction in mean depression score. Some patients experienced adverse effects, yet, they were temporary and alleviated within an hour. Moreover, the reduction in mean depression score remained at the end of the two weeks and during the follow up a month after. In summary, this study provides clinical evidence similar to Gutierrez et al.'s study, demonstrating the instant and long-term effectiveness of low-dose injections of ketamine without inducing intolerable, long-lasting side effects.

Sos et al.'s double-blind clinical trial investigating the efficacy of a single ketamine infusion in 27 hospitalised patients diagnosed with major depressive disorder. Participants received one ketamine and one placebo treatment session, order was randomised, with a one week interval between the two treatments. Patients were given ketamine infusions via their forearm and efficacy of treatment was measured using a clinician-administered scale, assessing severity of depression. Results found that those receiving ketamine experienced an alleviation in mood ratings during the week following with maximum effect on the seventh day. Overall, patients experienced no serious adverse effects during the period of study. Hence, ketamine has antidepressant and psychedelic effects that help alleviate the depressive symptoms in depressive patients in a clinical setting and is generally well-tolerated with minimal adverse effects, preventing patients' quality of life from deteriorating.

Overall, the findings from Gutierrez et al., Mandal et al., and Sos et al. collectively show that low doses of ketamine produces quick and clinically significant reduction in depressive symptoms when it is injected or infused into depressive patients. Across the three studies, ketamine was well-tolerated, with adverse effects experienced by the patients being short, mild and undistruptive to their quality of life. Thus, ketamine shows promise in providing a fast-acting alternative treatment to antidepressants which often take weeks for improvement to be seen. Even though ketamine-assisted depression treatment may show positive outcomes, ketamine can have negative effects such as disorientation, accidents, organ damage, cognitive impairment, and psychological dependence when used inappropriately. Although some suggest that the likelihood of addiction to ketamine is low (Campbell and Lobo; Ingrosso et al.), ketamine blocks NMDA receptors in the brain which increases the level of dopamine in areas of the brain linked to reward (Liu et al.). Due to this feeling of reward and pleasurable effect created by ketamine, the drug could easily be abused. This emphasises the importance of clinically supervised use of ketamine in depression treatment to ensure patients receive effective treatment with minimal risk of misuse and dependence.

Similar to ketamine's rapid antidepressant relief in clinical settings, psilocybin, a psychedelic that has a similar mechanism as mescaline, shows promising results in offering depressed individuals with more sustainable therapeutic benefits.

Psilocybin

Psilocybin, also known as the magic mushroom, is a classic psychedelic used often in religious, healing rituals but has gained attention as a potential treatment for depression. There are over 100 species of psychedelic, psilocybin-containing mushrooms currently known and most of these mushrooms belong to the genus *psilocybe* (Stamets). When psilocybin is consumed, it is converted into psilocin, the active form of psilocybin (Vollenweider and Smallridge). The chemical structure of psilocin is similar to neuroregulators such as serotonin, thus allowing psilocin to mimic serotonin and bind to serotonin receptors. Psilocin is an agonist that targets the serotonin system, therefore it amplifies the effect of serotonin by binding to a serotonin receptor (5-HT_{2A}) and activating them.

The biological mechanism of psilocybin's therapeutic effect has been difficult to identify. However, studies utilising functional MRI scans show the most promising, consistent and significant results in explaining the mechanism (Carhart-Harris et al.). Through the scans, it was found that psilocybin's psychedelic effect was linked to the reduction in blood flow and level of oxygen in venous blood in certain parts of the brain, the Default Mode Network (DMN) being the main area affected (Carhart-Harris et al.). This region is responsible for "self-reflection, emotional processing, social interaction, and mental exploration" (Azarias et al.). Therefore, when blood flow and oxygen level lower in these regions, there is an increase in neural connectivity, caused by the DMN's loss of control, which results in an unfixed self-identity (ego-dissolution), and because new brain networks are formed, communication becomes more free and brain is more neuroplastic. In terms of therapy, the flexibility in communication and thinking patterns as a result of psilocybin (psilocin), helps depressed individuals to break the depressive cycle and reduce depressive symptoms (Gattuso et al.).

Carhart-Harris et al.'s clinical open-label trial investigated the efficacy of psilocybin as a treatment for Treatment-Resistant Major Depression (TRMD) by assessing depressive symptoms in twenty patients, a week to six months post treatment. Two oral doses of psilocybin were given to patients, seven days apart and patients had to fill in a questionnaire to assess depressive symptoms. Overall, the treatment was well tolerated and no serious side effects were shown by patients. For the first five weeks of the treatment, depressive symptoms were reduced and results remained positive three to five months post-treatment. Therefore, psilocybin is promising for individuals with depression, especially TRMD, due to its fast antidepressant effects and long-term benefits such as the protection from relapse.

A double blind study conducted with similar intention as Carhart-Harris et al., assessing the efficacy and safety of psilocybin in adults diagnosed with Major Depressive Disorder (MDD), utilised a comparison group where participants were given niacin (Raison et al.). 104 participants were randomly assigned to either a 25mg of psilocybin or 100mg of niacin (placebo)

condition. The efficacy of the treatment was assessed using a clinician-rated questionnaire from baseline to day 43. Results found that the psilocybin condition showed a greater reduction in depressive symptoms and functional disability from baseline to day 43 of treatment, with minimal serious adverse effects. Raison et al. concluded that psilocybin can produce fast and significant reduction in depressive symptoms within days and is relatively durable as the treatment can last over a month.

Davis et al. investigated the effect of psilocybin therapy in MDD patients by utilising a randomised clinical trial, assigning 24 participants to immediate treatment condition group or delayed treatment condition group (8 week delay). The participants consisted of adults between the ages 21 to 75 years old with a diagnosis of MDD and have no history of antidepressant use. Psilocybin was given in gelatin capsules in two sessions and severity of depression was measured using a clinician-administered scale. Data showed that the psilocybin treatment was tolerable as the completion rate was 89% with minimal adverse effects, and participants experienced a reduction in depression severity from the day after the first session. Moreover, 71% of participants continued to respond to the treatment 4 weeks after the treatment. Davis et al. concluded that psilocybin therapy was safe, tolerable, and produced quick, sustainable reduction in depressive symptoms.

Psilocybin can be concluded to be an effective psychedelic in treating various types of depression such as TRMD and MDD. Unlike the studies done for mescaline-assisted depression treatment, these studies provide data on psilocybin when used to treat depression in a clinical setting with high levels of reliability from the standardised procedure (dosages, usage and timings). Within the three studies, some compare the efficacy of psilocybin by utilising a placebo, allowing researchers to isolate the effect of the drug which increases the overall internal validity of the results, ensuring that the change in severity of depression are a result of the drug taken instead of extraneous variables. The drug has also been shown to have long-term durability (up to 4 weeks) after few treatment sessions, which is longer than ketamine, suggesting that psilocybin-assisted depression treatment is sustainable. Moreover, psilocybin did not cause extreme adverse effects in participants and was associated with little abuse potential, highlighting the safety of the treatment. However, the studies all share a key limitation. They rely heavily on a small and selective sample, ranging from 20 to over 100 strictly screened and selected participants. This limits the generalisability of the findings on the effectiveness of psilocybin in treating depression, as the diversity of depressed individuals is not captured fully, thus emphasising the need for more representative trials.

Conclusion

When researching the three psychedelics' efficacy in treating depression, evidence highlights the substantial amount of research surrounding psilocybin's efficacy in treating depression through clinical trials that utilise standardised measures and highly controlled conditions. However, clinical studies like Carhart-Harris et al. and Sos et al. utilise randomised control trials (RCTs), where participants are split to receive either a placebo or treatment drug.

While RCTs allow researchers to easily compare the efficacy of drugs like mescaline, ketamine, and psilocybin, they raise ethical concerns. Concealing a potentially beneficial and effective treatment from participants in the placebo conditions may cause or worsen depression. Therefore, clinical trials on efficacy of psychedelics must be carefully done in a safe, careful manner to ensure studies are reliable yet ethical. Mescaline, on the other hand, has comparatively low numbers of research done, therefore, more research on mescaline as a treatment option for depression, especially in a clinical setting, needs to be done to provide depressive individuals a variety of psychedelic-assisted treatment options. Studies done on mescaline, tend to utilise questionnaires to assess the efficacy of the drug. This self-report method can potentially lead to the participants' display of demand characteristics, socially desirable response, subjectivity, and memory inaccuracies which can introduce under or overestimation of depression, reducing the overall validity and accuracy of the data.

In all three psychedelics, depressive participants consistently showed rapid improvements in their depressive symptoms without experiencing severe adverse effects. Mescaline has been studied in a more naturalistic setting and shows effectiveness and safety in unsupervised context, hence mescaline shows potential in treating depression in individuals that are unable to access treatment in a supervised, clinical setting. Although ketamine may require more supervision to minimise dependence, it also demonstrates antidepressant effects and maintains effect for up to two weeks. However, through this exploration, psilocybin could be considered the most effective psychedelic in treating depression due to its long-term efficacy, rapid improvement, minimal adverse effect and lack of abuse potential.

While psychedelics are relatively safe, depressive individuals with schizophrenia or psychosis, are advised not to take psychedelics as they can induce psychotic symptoms such as hallucinations, delusions, distorted perception, thus worsening the existing symptoms and psychosis. As psychedelics mimic the hallucination, vulnerable schizophrenic or psychotic individuals can experience relapse, episodes with greater severity, or long-lasting harm (Maćkowiak). For these reasons, clinical trials investigating the efficacy of psychedelics in treating depression exclude psychotic participants. Therefore, psychedelics can not be used safely for all clinical populations.

This literature review set out to explore the effectiveness of psychedelics, specifically mescaline, ketamine, and psilocybin, as an emerging tolerable treatment that alleviates depressive symptoms. Initially examining the limitations of conventional antidepressants and psychotherapy, this discussion highlighted the reason behind the growing interest and research in psychedelic-assisted therapy. Through clinical trials and naturalistic, unsupervised studies, it is evident that ketamine offers rapid relief, and psilocybin provides long-lasting, sustainable antidepressant effects, and mescaline, though less studied, shows reduction in depressive symptoms in long-term users in an unsupervised settings. Collectively, the findings from these studies suggest that psychedelics have potential to treat depression, especially treatment-resistant depression. Although psilocybin appears the most promising, more research must be done on a larger scale and more rigorously to ensure all three substances are safe, efficient, sustainable, and

their mechanisms are known. Therefore, this review reinforces the importance of additional and continued investigation surrounding psychedelic-assisted therapy as a practical depression treatment.

Works Cited

- Agin-Liebes, Gabrielle, et al. 'Naturalistic Use of Mescaline Is Associated with Self-Reported Psychiatric Improvements and Enduring Positive Life Changes'. *ACS Pharmacology & Translational Science*, vol. 4, no. 2, Apr. 2021, pp. 543–52. DOI.org (Crossref), <https://doi.org/10.1021/acspsci.1c00018>.
- Artigas, Francesc, et al. 'Mechanism of Action of Antidepressants'. *PSYCHOPHARMACOLOGY BULLETIN*, vol. 36, 2002.
- Artin, Hewa, et al. 'How Do Serotonergic Psychedelics Treat Depression: The Potential Role of Neuroplasticity'. *World Journal of Psychiatry*, vol. 11, no. 6, June 2021, pp. 201–14. DOI.org (Crossref), <https://doi.org/10.5498/wjp.v11.i6.201>.
- Azarias, Felipe Rici, et al. 'The Journey of the Default Mode Network: Development, Function, and Impact on Mental Health'. *Biology*, vol. 14, no. 4, Apr. 2025, p. 395. DOI.org (Crossref), <https://doi.org/10.3390/biology14040395>.
- Belouin, Sean J., and Jack E. Henningfield. 'Psychedelics: Where We Are Now, Why We Got Here, What We Must Do'. *Neuropharmacology*, vol. 142, Nov. 2018, pp. 7–19. DOI.org (Crossref), <https://doi.org/10.1016/j.neuropharm.2018.02.018>.
- Braund, Taylor A., et al. 'Antidepressant Side Effects and Their Impact on Treatment Outcome in People with Major Depressive Disorder: An iSPOT-D Report'. *Translational Psychiatry*, vol. 11, no. 1, Aug. 2021, p. 417. DOI.org (Crossref), <https://doi.org/10.1038/s41398-021-01533-1>.
- Breeksema, Joost J., et al. 'Patient Perspectives and Experiences with Psilocybin Treatment for Treatment-Resistant Depression: A Qualitative Study'. *Scientific Reports*, vol. 14, no. 1, Feb. 2024, p. 2929. DOI.org (Crossref), <https://doi.org/10.1038/s41598-024-53188-9>.
- Campbell, Rianne, and Mary Kay Lobo. 'A Short Burst of Reward Curbs the Addictiveness of Ketamine'. *Nature*, vol. 608, no. 7922, Aug. 2022, pp. 271–72. PubMed, <https://doi.org/10.1038/d41586-022-01948-w>.
- Carhart-Harris, R. L., et al. 'Psilocybin with Psychological Support for Treatment-Resistant Depression: Six-Month Follow-Up'. *Psychopharmacology*, vol. 235, no. 2, Feb. 2018, pp. 399–408. DOI.org (Crossref), <https://doi.org/10.1007/s00213-017-4771-x>.
- Carhart-Harris, Robin L., et al. 'Neural Correlates of the Psychedelic State as Determined by fMRI Studies with Psilocybin'. *Proceedings of the National Academy of Sciences*, vol. 109, no. 6, Feb. 2012, pp. 2138–43. DOI.org (Crossref), <https://doi.org/10.1073/pnas.1119598109>.
- Davis, Alan K., et al. 'Effects of Psilocybin-Assisted Therapy on Major Depressive Disorder: A Randomized Clinical Trial'. *JAMA Psychiatry*, vol. 78, no. 5, May 2021, p. 481. DOI.org (Crossref), <https://doi.org/10.1001/jamapsychiatry.2020.3285>.
- Davis, M. 'Mescaline: Excitatory Effects on Acoustic Startle Are Blocked by Serotonin2 Antagonists'. *Psychopharmacology*, vol. 93, no. 3, Nov. 1987. DOI.org (Crossref), <https://doi.org/10.1007/BF00187244>.

- Doblin, Richard E., et al. 'The Past and Future of Psychedelic Science: An Introduction to This Issue'. *Journal of Psychoactive Drugs*, vol. 51, no. 2, Mar. 2019, pp. 93–97. DOI.org (Crossref), <https://doi.org/10.1080/02791072.2019.1606472>.
- Fava, Giovanni A. 'May Antidepressant Drugs Worsen the Conditions They Are Supposed to Treat? The Clinical Foundations of the Oppositional Model of Tolerance'. *Therapeutic Advances in Psychopharmacology*, vol. 10, Jan. 2020, p. 2045125320970325. DOI.org (Crossref), <https://doi.org/10.1177/2045125320970325>.
- Gattuso, James J, et al. 'Default Mode Network Modulation by Psychedelics: A Systematic Review'. *International Journal of Neuropsychopharmacology*, vol. 26, no. 3, Mar. 2023, pp. 155–88. DOI.org (Crossref), <https://doi.org/10.1093/ijnp/pyac074>.
- Gaynes, Bradley N., et al. 'Defining Treatment-resistant Depression'. *Depression and Anxiety*, vol. 37, no. 2, Feb. 2020, pp. 134–45. DOI.org (Crossref), <https://doi.org/10.1002/da.22968>.
- "GBD Results." Institute for Health Metrics and Evaluation, vizhub.healthdata.org/gbd-results.
- Gutierrez, Gilmar, et al. 'IV Low Dose Ketamine Infusions for Treatment Resistant Depression: Results from a Five-Year Study at a Free Public Clinic in an Academic Hospital'. *Psychiatry Research*, vol. 335, May 2024, p. 115865. DOI.org (Crossref), <https://doi.org/10.1016/j.psychres.2024.115865>.
- Guttman, Erich. 'Artificial Psychoses Produced by Mescaline'. *Journal of Mental Science*, vol. 82, no. 338, May 1936, pp. 203–21. DOI.org (Crossref), <https://doi.org/10.1192/bjp.82.338.203>.
- Halpern, John H., et al. 'Psychological and Cognitive Effects of Long-Term Peyote Use Among Native Americans'. *Biological Psychiatry*, vol. 58, no. 8, Oct. 2005, pp. 624–31. DOI.org (Crossref), <https://doi.org/10.1016/j.biopsych.2005.06.038>.
- Husain, Muhammad Ishrat, et al. 'Serotonergic Psychedelics for Depression: What Do We Know about Neurobiological Mechanisms of Action?' *Frontiers in Psychiatry*, vol. 13, 2022, p. 1076459. PubMed, <https://doi.org/10.3389/fpsy.2022.1076459>.
- Ingrasso, Gianmarco, et al. 'Is There a Risk of Addiction to Ketamine during the Treatment of Depression? A Systematic Review of Available Literature'. *Journal of Psychopharmacology*, vol. 39, no. 1, Jan. 2025, pp. 49–65. DOI.org (Crossref), <https://doi.org/10.1177/02698811241303597>.
- Jay, Mike. *Mescaline: A Global History of the First Psychedelic*. Yale university press, 2019.
- Kadriu, Bashkim, et al. 'Ketamine and Serotonergic Psychedelics: Common Mechanisms Underlying the Effects of Rapid-Acting Antidepressants'. *International Journal of Neuropsychopharmacology*, vol. 24, no. 1, Jan. 2021, pp. 8–21. DOI.org (Crossref), <https://doi.org/10.1093/ijnp/pyaa087>.
- Kalfas, Michail, et al. 'Psychedelics for Treatment Resistant Depression: Are They Game Changers?' *Expert Opinion on Pharmacotherapy*, vol. 24, no. 18, Dec. 2023, pp. 2117–32. DOI.org (Crossref), <https://doi.org/10.1080/14656566.2023.2281582>.

- Kim, Jeongseop, et al. 'The Role of Glutamate Underlying Treatment-Resistant Depression'. *Clinical Psychopharmacology and Neuroscience: The Official Scientific Journal of the Korean College of Neuropsychopharmacology*, vol. 21, no. 3, Aug. 2023, pp. 429–46. PubMed, <https://doi.org/10.9758/cpn.22.1034>.
- Ko, Kwonmok, et al. 'Psychedelic Therapy for Depressive Symptoms: A Systematic Review and Meta-Analysis'. *Journal of Affective Disorders*, vol. 322, Feb. 2023, pp. 194–204. DOI.org (Crossref), <https://doi.org/10.1016/j.jad.2022.09.168>.
- Kyzar, Evan J., et al. 'Psychedelic Drugs in Biomedicine'. *Trends in Pharmacological Sciences*, vol. 38, no. 11, Nov. 2017, pp. 992–1005. DOI.org (Crossref), <https://doi.org/10.1016/j.tips.2017.08.003>.
- Liu, Yu, et al. 'Ketamine Abuse Potential and Use Disorder'. *Brain Research Bulletin*, vol. 126, Sept. 2016, pp. 68–73. DOI.org (Crossref), <https://doi.org/10.1016/j.brainresbull.2016.05.016>.
- Maćkowiak, Marzena. 'Psychedelics Action and Schizophrenia'. *Pharmacological Reports: PR*, vol. 75, no. 6, Dec. 2023, pp. 1350–61. PubMed, <https://doi.org/10.1007/s43440-023-00546-5>.
- Mandal, Suprio, et al. 'Efficacy of Ketamine Therapy in the Treatment of Depression'. *Indian Journal of Psychiatry*, vol. 61, no. 5, 2019, p. 480. DOI.org (Crossref), https://doi.org/10.4103/psychiatry.IndianJPsychiatry_484_18.
- McIntyre, Roger S., et al. 'Treatment-Resistant Depression: Definition, Prevalence, Detection, Management, and Investigational Interventions'. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, vol. 22, no. 3, Oct. 2023, pp. 394–412. PubMed, <https://doi.org/10.1002/wps.21120>.
- Muir, William W., and Richard A. Sams. 'Pharmacologic Principles and Pain: Pharmacokinetics and Pharmacodynamics'. *Handbook of Veterinary Pain Management*, Elsevier, 2009, pp. 113–40. DOI.org (Crossref), <https://doi.org/10.1016/B978-032304679-4.10007-3>.
- Raison, Charles L., et al. 'Single-Dose Psilocybin Treatment for Major Depressive Disorder: A Randomized Clinical Trial'. *JAMA*, vol. 330, no. 9, Sept. 2023, p. 843. DOI.org (Crossref), <https://doi.org/10.1001/jama.2023.14530>.
- Ramic, Enisa, et al. 'Assessment of the Antidepressant Side Effects Occurrence in Patients Treated in Primary Care'. *Materia Socio-Medica*, vol. 32, no. 2, June 2020, pp. 131–34. PubMed, <https://doi.org/10.5455/msm.2020.32.131-134>.
- Rogers, Carl R. "'Client-Centered' Psychotherapy." *Scientific American*, vol. 187, no. 5, 1952, pp. 66–75. JSTOR, <http://www.jstor.org/stable/24944053>. Accessed 6 Sept. 2025.
- Romeo, Bruno, et al. 'Efficacy of Psychedelic Treatments on Depressive Symptoms: A Meta-Analysis'. *Journal of Psychopharmacology*, vol. 34, no. 10, Oct. 2020, pp. 1079–85. DOI.org (Crossref), <https://doi.org/10.1177/0269881120919957>.
- Rondón Bernard, José Eduardo. 'Depression: A Review of Its Definition'. *MOJ Addiction Medicine & Therapy*, vol. 5, no. 1, Jan. 2018. DOI.org (Crossref), <https://doi.org/10.15406/mojamt.2018.05.00082>.

- Savalia, Neil K., et al. 'A Dendrite-Focused Framework for Understanding the Actions of Ketamine and Psychedelics'. *Trends in Neurosciences*, vol. 44, no. 4, Apr. 2021, pp. 260–75. DOI.org (Crossref), <https://doi.org/10.1016/j.tins.2020.11.008>.
- Sos, Peter, et al. 'Relationship of Ketamine's Antidepressant and Psychotomimetic Effects in Unipolar Depression'. *Neuro Endocrinology Letters*, vol. 34, no. 4, 2013, pp. 287–93.
- Stamets, Paul. *Psilocybin Mushrooms of the World: An Identification Guide*. Ten Speed Press, 1996.
- Stork, C.M., and S.M. Schreffler. 'Peyote'. *Encyclopedia of Toxicology*, Elsevier, 2014, pp. 841–43. DOI.org (Crossref), <https://doi.org/10.1016/B978-0-12-386454-3.00765-X>.
- Uthaug, Malin Vedøy, et al. 'The Epidemiology of Mescaline Use: Pattern of Use, Motivations for Consumption, and Perceived Consequences, Benefits, and Acute and Enduring Subjective Effects'. *Journal of Psychopharmacology*, vol. 36, no. 3, Mar. 2022, pp. 309–20. DOI.org (Crossref), <https://doi.org/10.1177/02698811211013583>.
- Vamvakopoulou, Ioanna A., et al. 'Mescaline: The Forgotten Psychedelic'. *Neuropharmacology*, vol. 222, Jan. 2023, p. 109294. DOI.org (Crossref), <https://doi.org/10.1016/j.neuropharm.2022.109294>.
- Vollenweider, Franz X., and John W. Smallridge. 'Classic Psychedelic Drugs: Update on Biological Mechanisms'. *Pharmacopsychiatry*, vol. 55, no. 03, May 2022, pp. 121–38. DOI.org (Crossref), <https://doi.org/10.1055/a-1721-2914>.
- Zanos, P., and T. D. Gould. 'Mechanisms of Ketamine Action as an Antidepressant'. *Molecular Psychiatry*, vol. 23, no. 4, Apr. 2018, pp. 801–11. PubMed, <https://doi.org/10.1038/mp.2017.255>.

Is Objectivity All in the Mind? Investigating the Psychological Construction of Health

By Kiyan Kapur

Abstract

Health indicators — such as cortisol, immune indicators, and heart rate variability — are usually considered as objective measurements of the internal physiological state of the body. However, this assumption has recently been challenged by psychological and neuroscientific research. This paper argues that health is not merely a reflection of biological facts, but is actively shaped by mental processes such as belief, bias, and emotional appraisal, drawing on evidence from cognitive psychology and affective neuroscience. Expectation, stress, mood, and interpretation have a significant influence on physiological function, and therefore, demand a reevaluation of objectivity in the health domain.

Research Question

To what extent do psychological processes such as cognitive bias and emotional interpretation shape physiological health measures typically considered objective?

Introduction

Biomarkers — such as heart rate variability, cortisol, and blood pressure — have historically been viewed in medical practice and psychological science as objective measures of biological function. The reason is that they are valid indicators that provide values that can be compared between individuals and groups, that can be quantified and replicated. For instance, a blood pressure reading of 120/80 mmHg is often categorized as 'normal'. Normal blood pressure thresholds are socially constructed cut-offs that create the impression of neutrality and uniformity. As a result, it has been perceived that these measures provide unambiguous information about the body's processes, which are separate from subjective perception.

However, this argument presupposes a mind-body dualism. Whereas, more recent work in psychology and neuroscience has led us to consider a relationship that is much more deeply embedded. Cognitive functions, such as belief, emotion, and perception, can influence biological systems in measurable ways. That influence extends to the interpretation of symptoms, as well as the regulation of physiological processes themselves.

This work has looked at the interaction in two parts. The first examines cognitive biases, including the placebo effect, confirmation bias, and optimism bias, to elucidate the role that expectations play in symptom interpretation and physiological response. The second addresses affective and psychophysiological processes that reshape biological systems, including stress, gene expression, and immunological activity.

The study begins with psychological expectations and mental construction, which determine and are responsible for the perception of health on a physiological level. Just as physiology and psychology are linked in a feedback relationship, cognitive psychology has described several biases and interpretative habits that influence people's relationships with their

bodily states. This section provides an overview of five mechanisms: the placebo effect, confirmation bias, optimism bias, cognitive dissonance, and cognitive distortions.

Placebo Effect and Expectation Bias

The placebo effect is the classical example of how psychological expectation can generate measurable biological effects. Simply put, it is defined by the fact that inert treatments or drugs appear to provide therapeutic benefits solely because the patient believes they do. This belief is thought to activate a specific neurochemical pathway (Benedetti, 2014) and therefore alter physiological processes that are typically disengaged from mental states.

Fabrizio Benedetti's research is central to the explanation of this phenomenon¹. His work describes how the expectation of pain relief alone could potentially activate the brain's endogenous opioid system. Volunteers were given a placebo but told they were being given a potent analgesic, which showed increased release of endorphins, which are neurochemicals that bind to opioid receptors to reduce pain perception. These changes were even blocked when naloxone, an opioid antagonist, was administered. Hence, this proved the body's internal pain-relief system could be pharmacologically triggered by belief alone.

Neuroimaging research further supports this link. Wager et al. used functional magnetic resonance imaging (fMRI) to demonstrate that placebo analgesia activates the prefrontal cortex². This structure subsequently modulates subcortical structures, such as the periaqueductal gray (PAG), which is involved in pain inhibition. Therefore, it suggests that expectancy not only influences self-reported measures of pain but also reorganizes the neural systems involved in sensory and affective components of pain processing. Additionally, placebo responses have been reported in depression and Parkinson's disease, with dopaminergic systems being activated by expectancy cues³. Such findings question the notion that the physiological response is an artifact of chemical agents or therapy. Instead, the belief that one is receiving care becomes a biologically active component of treatment.

The placebo effect shows that belief can produce measurable outcomes, such as reduced inflammation or altered neural activity, directly linking subjective expectation to objective biological change. This challenges the assumption that medical objectivity belongs exclusively to drug-based interventions since the mind itself can mimic the drug's physiological effects. Suppose belief alone can generate the same measurable outcomes. In that case, the definition of medical objectivity itself becomes unsettled: it must account not only for chemical causation, but also for the biological power of expectation.

The placebo effect thus constitutes compelling evidence that mind states are causal determinants of physiological effects. Subjective experiences like beliefs, and expectations

¹ Benedetti F. Placebo Effects: From Neurobiological Mechanisms to Clinical Applications. *Physiol Rev*. 2014;94(4):1207-1246.

² Wager TD, Rilling JK, Smith EE, et al. Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain. *Science*. 2004;303(5661):1162-1167.

³ de la Fuente-Fernández R, Ruth TJ, Sossi V, Schulzer M, Calne DB, Stoessl AJ. Expectation and Dopamine Release: Mechanism of the Placebo Effect in Parkinson's Disease. *Science*. 2001;293(5532):1164-1166.

generate measurable changes—this suggests that objectivity in health measurement is influenced by psychological states, and cannot be fully understood without accounting for them.

Confirmation Bias and Perceptual Distortion

Confirmation bias is the tendency to seek and interpret information in a manner that confirms one's pre-existing notions. In the healthcare context, confirmation bias has a significant impact on the detection, interpretation, and response to symptoms. Unlike with overt delusions, confirmation bias operates more subtly, influencing an individual's focus and assessment of physiological signs in a manner that tends to exacerbate pre-existing health anxieties or dismissals.

Although this psychological effect appears across a wide range of contexts, its relevance to health psychology is particularly significant. For instance, a person who is very confident of being ill is likely to begin attributing normal feelings—e.g., fatigue, muscular ache, or digestive discomfort—to signs of a previously unidentified illness. On the other hand, a person who is very confident about their health condition may tend to overlook symptoms that require medical evaluation. This selective attention has a direct influence on reporting behavior and a direct impact on physiological stress responses. Empirical studies have shown that expectancy and belief systems affect stress reactivity and immune function. For example, individuals who appraise a situation as threatening, even without objective danger, can experience higher levels of cortisol and increased levels of sympathetic arousal. These physiological changes are then linked to processes such as digestion, inflammation, and cardiovascular function, which then have a cascade of somatic effects that ultimately link back to cognitive processes. The bias, therefore, becomes self-perpetuating: the more people believe in the 'illness', the more the stress response is activated. Eventually, this creates physiological changes that seem to justify the original belief.

Furthermore, diagnostic delay is often the result of confirmation bias. People with dogmatic belief systems frequently reject new information, such as unfavorable test results or alternative explanations for events, if it conflicts with their existing beliefs. Furthermore, the information that a patient elects to disclose or not disclose, based on their faith or fear, can have a significant impact on the diagnostic process. In this way, the tests ordered and the data compiled are already filtered through a subjective human lens before any so-called 'objective' results are obtained. The denial of new information simply exacerbates the problem, making clinical communication even more challenging and resulting in unnecessary testing or inadequate care. Even with proper diagnostic instruments, perceptions remain clouded by hidden assumptions.

Consequently, confirmation bias functions both as a mental filter and a biological amplifier. When the body of a person shows traces of sickness, but the mind overlooks such signs, the question of objectivity comes into play: does it reside in the raw symptoms, or in the patient's lived awareness? In practice, medicine cannot disentangle the two.

Objectivity is not to be found either in symptoms or in the patient's experience of them.

Instead, it emerges at the point of convergence between subjective impressions and bodily evidence. It is there that human belief begins to influence biological life, and it suggests that what we call "objective data" is already intertwined with perception and interpretation.

Optimism Bias and the Underestimation of Health Risk

Optimism bias occurs when individuals underestimate their likelihood of experiencing adverse outcomes. Optimism bias can occur due to one's feeling well and being motivated, but it can also lead to people having dangerous and incorrect perceptions about their symptoms. In the context of disease prevention, this bias can lead to repeated underestimation of vulnerability, delay in response, and increased risk.

Sharot explained optimism bias as a widespread phenomenon observed among different groups, where people tend to underestimate their likelihood of contracting diseases, getting injured, or developing medical issues⁴. A change in risk perception has been seen even in high-risk settings. For example, smokers will generally accept the link between tobacco use and cancer, but at the same time, they will believe that they are less likely than other smokers to suffer an illness. Smokers reconstruct risk perception to support self-serving narratives. Therefore, prevention is an avoidance behaviour for smokers who visit the doctor's office for a check-up, change their lifestyle, or take part in early detection of symptoms.

The physiological consequences of this bias are not limited to the cognitive component of the process. Over-optimism can delay treatment for too long and allow pathological processes to become established. Although the individuals may feel healthy, subclinical vascular damage could be the result. Indeed, by the time the symptoms are impossible to ignore, the biological loss is much greater, and the efficacy of treatment is compromised. In that sense, cognitive underestimation has undeniable physiological consequences.

Additionally, optimism bias can also influence the way people report symptoms to healthcare professionals. Specifically, there is evidence that high dispositional optimists judge their symptoms as transient or mild⁵. Optimism bias may decrease anxiety related to health, but it may also prevent accurate reporting, which is needed for early identification and treatment. As such, it creates an imbalanced feedback mechanism: the greater an individual's underestimation of risk, the less likely they are to notice, interpret, or report corresponding symptoms, thereby increasing the possibility of disease progression. Thus, optimism bias demonstrates that humans are not entirely rational decision-makers, even in the face of serious health consequences, which challenges the Enlightenment ideal that objectivity leads to reasoned action. Instead, subjective emotion often overrules empirical truth.

Hence, optimism bias means more than a relentlessly optimistic approach to existence. It systematically warps the pathways through which health-relevant information is filtered, influencing cognitive measurement and physiological outcomes.

⁴ Sharot T. The optimism bias. *Curr Biol.* 2011;21(23):R941-R945.

⁵ Shepperd JA, Waters EA, Weinstein ND, Klein WM. A primer on unrealistic optimism. *Curr Dir Psychol Sci.* 2013;22(4):377-382

Cognitive Dissonance and Heuristics in Health Rationalization

Cognitive dissonance arises when individuals experience discomfort from holding opposing beliefs or engaging in an action that contradicts their values. To ease the pain, people often tend to make coping rationalizations. As an illustration, someone avoiding medical tests might say, 'No news is good news.' It feels reassuring, but it risks ignoring real health issues. Unfortunately, in medicine, dissonance is typically solved by altering one's reinterpretation rather than their harmful habits.

Festinger (1957) first proposed the idea of an individual's quest for internal consistency⁶. For example, a health-risk-alarmed smoker can cite anecdotal counterexamples—such as Jeanne Calment, who lived to be 122 years old while smoking—to imply that the risk has been overstated. This reinterpretation reduces psychological tension without altering the consequent physiological harm.

This type of thinking tends to rely on heuristics—mental shortcuts that aid the processing of complex information. These can be availability heuristics, such as utilizing vivid instances. Another heuristic is representativeness heuristics which entails forecasting patterns from surface resemblance. While useful for everyday decision-making, they have the potential to mislead health judgment. In the smoking example, the rare longevity case is treated as representative, despite contradicting large-scale epidemiological data. The consequence is continued engagement in risky behaviour based on a false sense of security.

Evidence suggests that unresolved dissonance has a biological impact. Low self-esteem participants who experienced performance failure, in the study by Pruessner et al. (1999)⁷, exhibited increased cortisol responses, indicating that psychological distress is translated into endocrine stress-related changes. The physiological consequences are not harmless; prolonged cortisol activation is linked to immunosuppression, cardiovascular stress, and premature aging.

Cognitive dissonance thus acts not just as a psychological moderator but also as a biological mediator. Dissonance resolution tends to prioritize self-concept over physical health, allowing individuals to suppress symptoms, delay medical consultations, or persist in self-destructive habits. In this way, dissonance and its cognitive offshoots reshape both perception and physiology, further challenging the separation of mental states from measurable health outcomes.

Cognitive Distortions and Symptom Interpretation

Cognitive distortions are automatic, erroneous thinking patterns that consistently skew perception. In health contexts, they can result in either heightened anxiety or greater complacency. They affect the perception of symptoms, acceptance of medical counsel, and the approach to healthcare, whether sought or evaded. In contrast to situational errors in Decision-making is often automatic and not always conscious, and therefore requires professional intervention to be recognized.

⁶ Festinger L. *A Theory of Cognitive Dissonance*. Stanford University Press; 1957.

⁷ Pruessner JC, Hellhammer DH, Kirschbaum C. Low self-esteem, induced failure and the adrenocortical stress response. *Pers Individ Dif*. 1999;27(3):477-489.

Other typical cognitive biases include catastrophizing, minimization, all-or-nothing thinking, and personalization. For example, an individual with a headache may think it is a sign of a brain tumor (catastrophizing). In cases such as health anxiety or hypochondriasis, this subjective sense of alarm can escalate into a clinically recognized disorder, where the anxious cognitive state itself becomes a source of measurable physiological stress. Moreover, a person with chronic fatigue may attribute it to overworking and ignoring possibly underlying health complications (minimization). Similarly, the patient may believe that if they cannot resume their complete exercise program immediately, their career is a failure and they may as well abandon rehabilitation altogether (all-or-nothing thinking). Others may interpret a delayed diagnosis as evidence that they are to blame for being sick or being punished somehow (personalization), reinforcing a cycle of shame and denial of symptoms. Harvard Health (2021) clarifies that such cognitions are linked with adverse health outcomes, since they distort the ability to make realistic assumptions about the symptoms and are sure to delay the appropriate medical attention⁸.

Additionally, distorted thinking may also contribute to maladaptive behaviours. Catastrophizers exhibit extreme sensitivity to health warnings. They consistently seek health-related comfort and reassurance regarding their non-threatening human symptoms, which can lead to over-diagnosis or excessive treatment⁹. On the other hand, minimizers do not seek help even when there are warning signals of dire consequences because they constantly downplay the significance of their symptoms. Both patterns create inconsistencies between the actual state of the body and its treatment, thereby affecting disease courses and treatment responses. These thoughts and beliefs have been associated in related literature on depression and anxiety¹⁰. These biological effects demonstrate that distorted cognition not only alters perception but also causally contributes to body dysfunction.

Distorted thinking is, therefore, a process by which the mind distorts what is being measured in the body, reinforcing the paper's central claim that psychological states and physiological outcomes are deeply co-constructed. While cognitive biases record how systems of belief influence perception and physiological consequences, their response does not explain the entire spectrum of mind-body interactions. The following level of this research takes into consideration the affective and neurological processes by which mood, emotion, and meaning-making control biological systems. These processes, unlike discrete cognitive distortions, function through integrated feedback loops that integrate psychological states with hormonal, immunological, and even genomic processes. Understanding this integration is essential to supporting the paper's central idea.

⁸ Harvard Health Publishing. Cognitive distortions: When your mind plays tricks on you. Harvard Health Blog. 2021. Available from: <https://www.health.harvard.edu/blog>

⁹ Gautreau CM, Sherry SB, Sherry DL, Birnie KA, Mackinnon SP, Stewart SH. Behavioural and Cognitive Psychotherapy. 2015;43(4):502–512.

¹⁰ Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of depression and anxiety on cardiovascular disease: Stress, autonomic dysregulation, and inflammation. *Neurosci Biobehav Rev*. 2012;36(1):187-192.

Allostatic Load and the Biological Cost of Stress

A core problem in treating health markers as “objective” is that they can be altered by demands that are not inherently dangerous. The body’s need to adapt to changing conditions — known as allostasis — ensures survival by reconfiguring physiological parameters for environmental demands. For example, during a high-stakes presentation, heart rate and cortisol levels may rise sharply to mobilize energy, an adaptive shift that helps performance. These reactions are protective in the short run.

However, the cumulative effect of repetitive activation of the same stress pathways causes wear and tear in the body, termed as allostatic load. This is a term introduced by McEwen that describes the insidious effects of chronic adaptation in multiple systems—ranging from immune suppression, which leaves the body vulnerable to infection, to metabolic disruption that can drive insulin resistance and cardiovascular risk¹¹. These may be measured in terms of biomarkers, such as elevated cortisol levels, heightened inflammatory cytokine levels, and reduced heart rate variability.

The neurobiological mechanism behind the process is the hypothalamic–pituitary–adrenal (HPA) axis. Activation of this cycle repeatedly, even in the absence of life-threatening situations, leads to chronic cortisol release. In the long term, this disrupts immune surveillance, alters glucose metabolism, and is associated with reduced hippocampal volume. Chronic psychological stress has been associated with shorter telomeres and lower telomerase activity in immune cells in a paper published by Epel et al¹². Sapolsky highlighted these findings in his commentary¹³.

Crucially, these physiological changes do not require objectively extreme stressors. Cognitive patterns such as catastrophizing, chronic pessimism, or learned threat responses can sustain high biomarker levels even in relatively safe environments. This raises a critical epistemological question: if such markers can be elevated purely through subjective interpretation of events, can they be treated as “objective” evidence of stress in research or diagnosis? Allostatic load suggests that without accounting for the psychological component, biomarker readings, if interpreted without considering psychological factors, may misrepresent an individual’s health status.

Importantly, the stress-producing causes of allostatic load are interpretive and subjective, not acute or life-threatening. For instance, role conflict, financial instability, or emotional invalidation has the potential to activate chronic physiological activation if they are subjectively interpreted as being uncontrollable or threatening. Therefore, research shows that the brain’s appraisal of threat, rather than the event itself, largely determines the magnitude of the stress response.”. What is perceived as manageable to one person can be internalized by another as detrimental—and it is this process of internalization that brings about biological effects.

This position implies that the intervention can occur through either of two routes: intervening with external stressors or changing the person’s appraisal of those stressors. When

¹¹ McEwen BS. Stress, adaptation, and disease: Allostasis and allostatic load. *Ann N Y Acad Sci.* 1998;840:33-44.

¹² Epel ES, Blackburn EH, Lin J, et al. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci USA.* 2004;101(49):17312-17315.

¹³ Sapolsky RM. *Why Zebras Don’t Get Ulcers.* 3rd ed. New York: Holt Paperbacks; 2004.

the stressor is concrete and adjustable, such as excessive work on the job or unsafe housing, addressing external pressures may offer the most immediate relief. However, in the vast majority of cases, individuals are not able to remove external pressures entirely. Therefore, it is possible that strategies with perception as the primary focus, such as cognitive reappraisal, mindfulness-based stress reduction, and resilience training, could be more effective, as they diminish both the physiological consequences of stress without completely removing all stressors. As perception is a key factor in causing allostatic load, these strategies may yield benefits that are broader and more durable.

The report that mood can alter gene expression marks a significant shift in the interaction of biology and psychology. For decades, gene function was supposed to be an immutable, inherited template that was only switched on or off, due to biological requirements or environmental constraints, instead of our mood. Current research in social genomics provides evidence that emotional states, particularly habitual or long-term well-being or distress, can directly influence which genes are active, and in turn, affect immune and inflammatory processes.

Fredrickson et al. reviewed the effects of two types of well-being on gene expression¹⁴. Hedonic well-being, which is based on feelings of pleasure and comfort, is anticipated to represent a different phenomenon compared to eudaimonic well-being, based on having purpose, meaning, and engagement in life. These states can be determined by using valid self-report questionnaires, along with genomic analyses that examine immune cell function. Individuals with high eudaimonic well-being exhibited lower expression of pro-inflammatory genes, resulting in a more pronounced antiviral response, whereas those with higher hedonic experiences displayed the opposite response. This suggests that not all forms of pleasure produce the same biological signature; instead, there are low-risk genetic profiles associated with more meaning-oriented welfare.

Gene expression in immune cells, particularly leukocytes, displays unique sensitivity to affective states. Chronic stress and negative affect are associated with increased activity of inflammation-associated genes (e.g., NF- κ B signalling), which are involved in many diseases, including depression and cardiovascular disease. Conversely, experiences of social connection, purpose, and emotional regulation tend to promote gene expression profiles associated with resilience and immune strength.

Hence, from the cumulative wear of allostatic load to the molecular shifts in gene expression, psychological states leave measurable and lasting marks on the body. Yet this raises an open question: if the mind shapes biomarkers, should their value rest on metaphysical objectivity, or is their predictive reliability for diagnosis and treatment enough?

Cognitive Reappraisal and Neural Regulation of Stress

One difficulty in treating stress-related biomarkers as objective is that they can shift depending on how an event is mentally framed. Cognitive reappraisal offers a solution to this

¹⁴ Fredrickson BL, Grewen KM, Coffey KA, et al. A functional genomic perspective on human well-being. *Proc Natl Acad Sci USA*. 2013;110(33):13684-13689.

problem.

Defined as the active reinterpretation of a potentially aversive encounter in a way that alters its emotional impact, it changes not just the felt experience but the body's physiological response. For instance, someone who views a tense meeting as an opportunity to contribute rather than a personal trial will likely exhibit reduced cortisol release and a steadier heart rate. Unlike denial or suppression, cognitive reappraisal is an adaptive strategy for emotion regulation, mediated by higher-order cognitive processes that affect stress-related neural circuits. Neuroscientific

Evidence has shown that top-down modulation can decrease amygdala activity and reduce HPA-axis activation; therefore, biological markers of stress are subject to cognitive interpretation as much as to external stress.

Ochsner and Gross had some of the strongest empirical evidence for this process. In a series of fMRI studies, they demonstrated that individuals who actively reappraised emotional stimuli caused increased activation in the prefrontal cortex (PFC) and decreased activation in the amygdala¹⁵. The PFC, which is responsible for executive function and cognitive control, has a top-down effect on limbic structures, such as the amygdala, which are involved in fear and stress responses. This brain region interaction reduces autonomic stimulation, including heart rate, blood pressure, and cortisol secretion.

Above all, cognitive reappraisal has been linked to improved physiological health in both clinical and non-clinical samples. Research shows that individuals who employ reappraisal strategies during stress often exhibit higher heart rate variability and reduced inflammatory activity^{16,17}.

Troy et al. add that reappraisal ability can buffer the effects of stress on depression¹⁸. For patients with social anxiety disorders, reappraisal-oriented cognitive behavioral therapy (CBT) has been found to increase prefrontal connectivity during emotional regulation, decreasing the severity of symptoms and reducing the amygdala reactions¹⁹.

These findings are robust, but there are a few caveats to their interpretation. Much of the Neuroscientific evidence is based on fMRI experiments that do not always provide a complete picture of how reappraisal works with real-life stressors, raising questions about ecological validity. Moreover, the more familiar relationship between amygdala exertion and control by the prefrontal cortex is a correlational one: it is an expression of the possibility of top-down control, but not of how to achieve it. Lastly, other biomarkers like inflammation and heart rate variability are susceptible to a multitude of factors, including nutrition, genetics, exercise, and sleep, and

¹⁵ Ochsner KN, Gross JJ. The cognitive control of emotion. *Trends Cogn Sci*. 2005;9(5):242-249.

¹⁶ Applehans BM, Luecken LJ. Heart rate variability and pain: Associations of two interrelated homeostatic processes. *Biol Psychol*. 2006;71(2):175-182.

¹⁷ Fabiansson EC, Denson TF, Moulds ML, Grisham JR. Don't look back in anger: Responsiveness to stressors and symptoms of depression. *Emotion*. 2012;12(6):1058-1069.

¹⁸ Troy AS, Wilhelm FH, Shallcross AJ, Mauss IB. Seeing the silver lining: Cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion*. 2010;10(6):783-795.

¹⁹ Goldin PR, Ziv M, Jazaieri H, Hahn K, Gross JJ. MBSR vs aerobic exercise in social anxiety: fMRI and symptom outcomes. *J Consult Clin Psychol*. 2014;82(5):927-940.

therefore any benefit associated with reuse may be indirect or incomplete. Best. However, research proves that even within the same body, there are numerous contributory factors to good physical foundations.

Meaning-Making and Biomarkers of Health

When cognitive function improves, the act of experiencing—through objectives, ideals, or knowledge of existence—mirrors physiologic outcomes in significant ways. Meaning-making has evolved beyond affective or philosophical thinking and is now described as a psychobiological process that regulates immune response, endocrine control, and general well-being. Individuals who find meaning and purpose in adverse circumstances typically have healthier biomarker profiles, suggesting that interpretation has a direct impact on internal bodily processes.

Davidson et al. also demonstrated that subjects who underwent an eight-week mindfulness course exhibited improved influenza antibody titers and altered brain activity compared to controls, thereby contributing to the evidence linking reflective practice with objective immune responses²⁰. These shifts in immune responses reflect a direct link between emotional awareness and control of the system.

Recently, Chida and Steptoe meta-analyzed over 70 studies that investigated the relationship between psychological well-being and several biomarkers, including interleukin-6, cortisol, and C-reactive protein²¹. The findings of their meta-analysis displayed that positive affect, particularly when it was based on important life events, was associated with lower inflammatory activity and better autonomic balance. This was true in healthy populations as well as in the presence of chronic diseases, thereby suggesting that meaning-making has health benefits independent of initial health status.

Importantly, not all stress is harmful. In a meaningful way, or as a consequence of the direct coping ability, a person will exhibit different biological responses to the crisis compared with what is known as eustress. Studies have shown that when people construe challenges as meaningful or as opportunities for personal growth, their physiological stress responses are more adaptive, resulting in faster recovery from cortisol levels and heart rate variability²². These are not merely semantic distinctions; they suggest that how an experience is framed determines whether stress will weaken or build health systems.

The ability to derive meaning from life's challenges serves as a physiologic modulator of the stress response, affecting the hormonal environment of the system and reducing immunological function. This means that physiological measures are irrevocably intertwined with the meaning of experience.

²⁰ Davidson RJ, Kabat-Zinn J, Schumacher J, et al. Alterations in Brain and Immune Function Produced by Mindfulness Meditation. *Psychosom Med.* 2003;65(4):564-570.

²¹ Chida Y, Steptoe A. Positive psychological well-being and mortality: A quantitative review of prospective observational studies. *Psychosom Med.* 2008;70(7):741-756.

²² Seery MD, Holman EA, Silver RC. Whatever does not kill us: Cumulative lifetime adversity, vulnerability, and resilience. *J Pers Soc Psychol.* 2010;99(6):1025-1041.

Neuroimmunological Integration and the Mind-Body Feedback Loop

Neuroimmunology has revolutionized our understanding of the relationship between the mind and body. Once considered completely different areas, immunology and neuroscience are now understood as a communication process that involves a two-way interaction. Immune signals have the power to change brain function and mood, and psychological states affect immune function.

The brain and immune system interact with each other through the use of similar signal molecules, including cytokines, neurotransmitters, and hormones. For example, pro-inflammatory cytokines such as interleukin-1 β and TNF- α , which accumulate during disease or injury, also impact mood, cognition, and energy²³. These molecules can inspire "behavior of the disease", such as fatigue, recurrence, and depression, reflecting the effect of Immunity activation on the mental state.

Furthermore, the brain can exert a top-down regulatory effect on the immune system. Emotional regulation in the prefrontal cortex influences the hypothalamus, which in turn adjusts the release of cortisol and other hormones involved in immune response. Positive affect, strong social Bonds and effective coping strategies are linked to enhanced resistance to infection, stronger vaccine responses, and faster wound healing²⁴. These benefits are not gained through pharmacological treatment but through psychological states that enhance defensive response.

Notably, the immune system can "remember" psychological states. Chronic stress has been associated with long-term changes in immune cell activity, including reduced activity of natural killer cells and increased inflammation²⁵. These effects can persist long after the stressor has passed, indicating that immune health is a reflection of the cumulative impact of mental life.

There is a substantial body of evidence supporting associations between psychological states and immune functioning, mediated by mechanisms including cytokine signaling, top-down neural regulation, and immune memory responses. Neuroimmunology is demonstrating that mental states can have potent, ongoing physiological effects. This presents a significant constraint for the medical notion of immunological markers as a (bio)objective health indicator. In reality, these responses were predicated not solely on the identification of pathogens but also on the unique experiences and fears associated with the sickness in their individual narratives.

Conclusion

This article has demonstrated that even physiological health measures, which are intended to assess bodily states, are influenced by psychological processes. Cognitive biases, such as placebo effects, confirmation biases, and optimism bias, are all examples that show how

²³ Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW. From inflammation to sickness and depression: When the immune system subjugates the brain. *Nat Rev Neurosci.* 2008;9(1):46-56.

²⁴ Kiecolt-Glaser JK, McGuire L, Robles TF, Glaser R. Emotions, morbidity, and mortality: New perspectives from psychoneuroimmunology. *Annu Rev Psychol.* 2002;53:83-107.

²⁵ Segerstrom SC, Miller GE. Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychol Bull.* 2004;130(4):601-630.

belief and expectation can not only affect how symptoms are perceived but also influence biological reactions. These findings are supported and extended by research in affective neuroscience and psychoneuroimmunology, which shows that mood, stress, and meaning-making actively regulate gene expression, immune function, and neural activity.

Across both cognitive and physiological domains, the evidence points to a common conclusion: health and expression cannot be accurately portrayed without reference to psychological types of mediation through expression. What appears as objective—cortisol levels, immune profiles, or even gene expression—is in many cases a reflection of the individual's emotional and cognitive landscape. These biological markers exist in a context that is co-created by the mind, and we can regard them as distinct. This does not mean that biology is irrelevant or illusory, but rather that it is incomplete without consideration of psychological understanding. The data to be discussed here will necessitate a dynamic model of health—namely, a context-dependent construction. In this view, the mind is not merely a passive observer of bodily states, but an active participant in their creation.

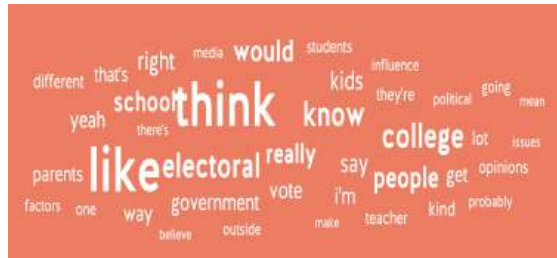
Interpreting objectivity as a socially constructed phenomenon encourages an integrated methodology towards health. This notion connects psychology with physical areas and recognizes that the mind is a principal agent in answering what is eventually accepted to be true.

Works Cited

1. Benedetti, Fabrizio. Placebo Effects: From Neurobiological Mechanisms to Clinical Applications. *Physiological Reviews*, vol. 94, no. 4, 2014, pp. 1207-1246.
2. Wager, Tor D., et al. "Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain." *Science*, vol. 303, no. 5661, 2004, pp. 1162-1167.
3. de la Fuente-Fernández, Raúl, et al. "Expectation and Dopamine Release: Mechanism of the Placebo Effect in Parkinson's Disease." *Science*, vol. 293, no. 5532, 2001, pp. 1164-1166.
4. Sharot, Tali. "The Optimism Bias." *Current Biology*, vol. 21, no. 23, 2011, pp. R941-R945.
5. Shepperd, James A., et al. "A Primer on Unrealistic Optimism." *Current Directions in Psychological Science*, vol. 22, no. 4, 2013, pp. 377-382.
6. Festinger, Leon. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.
7. Pruessner, Jens C., et al. "Low Self-Esteem, Induced Failure and the Adrenocortical Stress Response." *Personality and Individual Differences*, vol. 27, no. 3, 1999, pp. 477-489.
8. Harvard Health Publishing. "Cognitive Distortions: When Your Mind Plays Tricks on You." *Harvard Health Blog*, 2021, <https://www.health.harvard.edu/blog>.
9. Gautreau, Catherine M., et al. "Behavioural and Cognitive Psychotherapy." *Behavioural and Cognitive Psychotherapy*, vol. 43, no. 4, 2015, pp. 502-512.
10. Kemp, Andrew H., et al. "Impact of Depression and Anxiety on Cardiovascular Disease: Stress, Autonomic Dysregulation, and Inflammation." *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 1, 2012, pp. 187-192.
11. McEwen, Bruce S. "Stress, Adaptation, and Disease: Allostasis and Allostatic Load." *Annals of the New York Academy of Sciences*, vol. 840, 1998, pp. 33-44.
12. Epel, Elissa S., et al. "Accelerated Telomere Shortening in Response to Life Stress." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 49, 2004, pp. 17312-17315.
13. Sapolsky, Robert M. *Why Zebras Don't Get Ulcers*. 3rd ed., Holt Paperbacks, 2004.
14. Fredrickson, Barbara L., et al. "A Functional Genomic Perspective on Human Well-Being." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 33, 2013, pp. 13684-13689.
15. Ochsner, Kevin N., and James J. Gross. "The Cognitive Control of Emotion." *Trends in Cognitive Sciences*, vol. 9, no. 5, 2005, pp. 242-249.
16. Applehans, Bradley M., and Laurie J. Luecken. "Heart Rate Variability and Pain: Associations of Two Interrelated Homeostatic Processes." *Biological Psychology*, vol. 71, no. 2, 2006, pp. 175-182.
17. Fabiansson, Emma C., et al. "Don't Look Back in Anger: Responsiveness to Stressors and Symptoms of Depression." *Emotion*, vol. 12, no. 6, 2012, pp. 1058-1069.
18. Troy, Allison S., et al. "Seeing the Silver Lining: Cognitive Reappraisal Ability Moderates the Relationship between Stress and Depressive Symptoms." *Emotion*, vol. 10, no. 6, 2010, pp. 783-795.

19. Goldin, Philippe R., et al. "MBSR vs Aerobic Exercise in Social Anxiety: fMRI and Symptom Outcomes." *Journal of Consulting and Clinical Psychology*, vol. 82, no. 5, 2014, pp. 927-940.
20. Davidson, Richard J., et al. "Alterations in Brain and Immune Function Produced by Mindfulness Meditation." *Psychosomatic Medicine*, vol. 65, no. 4, 2003, pp. 564-570.
21. Chida, Yoichi, and Andrew Steptoe. "Positive Psychological Well-Being and Mortality: A Quantitative Review of Prospective Observational Studies." *Psychosomatic Medicine*, vol. 70, no. 7, 2008, pp. 741-756.
22. Seery, Mark D., et al. "Whatever Does Not Kill Us: Cumulative Lifetime Adversity, Vulnerability, and Resilience." *Journal of Personality and Social Psychology*, vol. 99, no. 6, 2010, pp. 1025-1041.
23. Dantzer, Robert, et al. "From Inflammation to Sickness and Depression: When the Immune System Subjugates the Brain." *Nature Reviews Neuroscience*, vol. 9, no. 1, 2008, pp. 46-56.
24. Kiecolt-Glaser, Janice K., et al. "Emotions, Morbidity, and Mortality: New Perspectives from Psychoneuroimmunology." *Annual Review of Psychology*, vol. 53, 2002, pp. 83-107.
25. Segerstrom, Suzanne C., and Gregory E. Miller. "Psychological Stress and the Human Immune System: A Meta-Analytic Study of 30 Years of Inquiry." *Psychological Bulletin*, vol. 130, no. 4, 2004, pp. 601-630.

Factors that Influence Youth Perceptions of the Electoral College By Tomer Fine



Abstract

The Electoral College has been a central but controversial feature of U.S. presidential elections since 1787, with growing calls for reform, particularly among younger generations. This study examines the factors that influence high school students' perceptions of the Electoral College, focusing on political activity, ideological affiliation, and sources of civic knowledge. Using a mixed-method approach, data were collected from 179 Calabasas High School students through surveys and from three government teachers through semi-structured interviews. Survey items drew from prior research, including Pew Research Center and Maffucci, while interviews looked into educators' perspectives on how youth learn about democratic institutions. Findings indicate that politically inactive youth rely more heavily on in-school education for information about the Electoral College, while politically active youth demonstrate higher confidence in their knowledge and greater familiarity with reform options. Political ideology strongly shaped opinions: conservative-leaning students tended to support retaining the Electoral College, while liberal-leaning students favored reforms such as the national popular vote. Contrary to expectations, teacher interviews revealed that outside-school influences (e.g., family, peers, and social media) outweigh classroom instruction in influencing student perceptions. These results show the growing importance of civic education and extracurricular political engagement in preparing the next generation to evaluate and reform democratic institutions.

I. Introduction

The Electoral College, a cornerstone of the U.S. presidential election system since its establishment in 1787, has long been a subject of debate and criticism. While it was originally designed as a compromise between electing the president by Congress and a popular vote, many Americans, particularly youth, now view the system as archaic and unrepresentative. Young people, who are becoming increasingly civically engaged, often feel disillusioned by the Electoral College, perceiving it as a system that undermines the democratic principle of "one person, one vote" (Hayden). Concerns about the Electoral College are particularly acute in non-swing states, which account for 82 percent of the U.S. population, where electoral results are typically predetermined (Chappell). This leaves many voters, especially youth, with the sense that their votes carry little weight. Given the growing civic awareness among young people, exploring the factors that influence their views on the Electoral College is essential for understanding their overall opinions toward electoral reform and democracy.

A study conducted at Calabasas High School (CHS), a diverse campus, provides a valuable opportunity to explore youth perceptions of the Electoral College. A recent Student Mock Election preceding the 2024 election, conducted at the high school through the California Secretary of State's office, revealed that approximately 50 percent of students identified as Democrats and 42.5 percent identified as Republicans. These results reflect similar trends observed in California as a whole, where 51 percent of those aged 14–18 identify as Democrats or lean Democrat, 25 percent identify as Republicans or lean Republican, and 23 percent have no political lean (Public Policy Institute of California). The CHS students' political diversity, combined with the school's representation of various socioeconomic and educational backgrounds, makes it an ideal setting to examine how factors such as political engagement, party affiliation, and civic education shape young people's attitudes toward the Electoral College. Understanding these perspectives is essential for addressing many young voters' discontent toward a system that often marginalizes non-swing states and devalues individual votes. Information from such a study could help promote electoral reforms that resonate with the democratic values of the next generation, creating greater trust and participation in the United States' political process.

II. Literature Review

This study aims to analyze factors influencing youth perceptions of the Electoral College and preferences for reform options that have become increasingly popular in recent decades. This inquiry stems from the growing phenomenon of youth-led political advocacy and reform movements, which raise questions about how young Americans view traditional democratic institutions such as the Electoral College. To refine their focus, the researcher drew on academic sources examining historical reform efforts, partisan divides over the Electoral College, and trends in youth civic participation.

History of Electoral College Reform Alternatives

Ever since the establishment of the Electoral College, reform movements have had long-lasting impacts on the past and present. They have rapidly grown in popularity over the last century due to numerous major factors, including the expansion of the electorate and unjust electoral outcomes (Rolat). University of Pennsylvania Law Professor Walter Clark provides an early and detailed review of the Electoral College's evolution during the first hundred years of U.S. history. His paper discusses how the founding fathers' original intentions for including the Electoral College in the Constitution made the system inherently corrupt, as it prevented the "common man" from having a major say in presidential elections. To resolve this concern, Clark proposed the adoption of a national popular vote, sparking a movement that would continue into the present day (Clark 738). Building on this foundation established by Clark, Whitaker et al. structured their paper around three alternatives to the Electoral College: the Proportional Plan, the District Plan, and the National Popular Vote (NPV). As defined by the study, the Proportional Plan recommends allocating a state's electoral votes in proportion to the popular vote each

candidate receives in that state. For example, if a candidate wins 60% of the vote in a state, they would receive 60% of that state's electoral votes. The District Plan, currently implemented in Maine and Nebraska, allocates one electoral vote to the winner of each congressional district; the remaining two votes are awarded to the candidate who wins the statewide popular vote. Lastly, the NPV guarantees the presidency to the winner of the national popular vote. The authors analyze these alternatives through case studies, particularly the contested 2000 election, to assess their practicality and fairness. They ultimately favor the NPV, recognizing that its ability to equalize voting power across states is essential to prevent scenarios where the Electoral College contradicts the popular vote (Whitaker).

Partisan Divide on the Electoral College

The divide over the Electoral College is heavily influenced by partisan perspectives (committed members of a political party), candidate preference, and perceptions of electoral fairness. A study by Aldrich et al. on the 2004 presidential election found that support for Electoral College reform was due to three main factors: candidate preference, perceptions of which candidate would benefit from the current system, and support for a popular vote system. For example, Democratic voters were more likely to favor reform if they believed the Republican candidate would win in their state, and vice versa (Aldrich et al.). Similarly, a study by Karp et al. using 2008 panel survey data revealed that Americans, in general, support reforms such as nationalizing elections or adopting a national popular vote, though Republicans were 10 percent less likely to favor such changes compared to Democrats. This displays a large partisan gap in overall support for electoral reform, with conservatives preferring to stick to more traditional systems and liberals being more open to new ideas (Karp et al.). Contrary to these previous studies, research on the 2020 election by Sances et al. found that voters whose preferred candidate lost were no more likely than others to question the legitimacy of the election process, including the Electoral College. The researcher analyzed 76 polls asking about the legitimacy of the 2000, 2016, and 2020 presidential elections. The polls were originally developed by 12 notable polling groups, including Quinnipiac University and Gallup polls. The findings of the study revealed that a large portion of the losing candidate's supporters did not believe there was some sort of electoral fraud in the 2000, 2016, and 2020 elections, regardless of political party affiliation (Sances et al.). This study substantiates the conclusion that support for differing political ideologies does not correlate with Electoral Reform preference, an idea refuted by Aldrich et al. and Karp et al.

Increased Youth Civic Participation in America

Youth civic participation has become a topic of growing interest, particularly in understanding what motivates young Americans to engage politically and how their engagement influences their perceptions of democratic institutions like the Electoral College. Fisher examines themes in youth political participation, focusing on the 2008 election, which saw one of the highest youth voter turnouts in U.S. history. It found that two-thirds of young voters supported

Barack Obama, emphasizing a Democratic lean among younger demographics due to different campaign tactics utilized by the Obama campaign, including “social media, messages of hope, and grassroots mobilization” (Fisher 135). Similarly, Syvertsen et al. analyzed data from *Monitoring the Future*, a long-term survey of high school seniors, to track trends in youth civic engagement since 1976. It revealed that while traditional political participation, such as voting, has decreased in some areas, recent “cohorts of young people are increasingly drawn to volunteerism and other forms of community action” (Syvertsen et al. 589). These findings offer important context for understanding youth involvement in politics. As younger generations are getting more involved in their community, their engagement with local politics and opinions about laws and policies is becoming more defined.

Foundational Sources

Two main foundational sources helped guide the path of this research study. The first foundational source was by The Pew Research Center, a non-partisan think tank, which conducted a study titled “Majority of Americans Continue to Favor Moving Away from Electoral College.” It found that nearly two-thirds of Americans support reforming the Electoral College to ensure the presidency is decided by the national popular vote. This can be broken down even further, as younger adults under 50 are most supportive of change, with 69% in favor, compared to only 58% of those aged 65 and older (Pew Research Center). Moreover, the study found that education, political party preference, and knowledge about politics as a whole all influenced overall perception. Similarly, a study by Maria Maffucci called “The Electoral College: A System for the People?” explores the perceptions of students regarding the Electoral College at the University of Maine, finding that most participants supported abolishing the Electoral College but lacked sufficient knowledge of its workings. The methodology used by this paper included an online survey of 20 questions regarding differing levels of political engagement among students and knowledge about existing reform options. Once the researcher collected over 400 responses from students at the university, they began to analyze and organize it into numerical data, including statistical graphs and charts (Maffucci). The researcher compared factors such as political ideology and political activity to the overall perception of the Electoral College and whether the system works efficiently. These studies served as the basis for the current study, with the researcher combining elements from both Maffucci and the Pew Research Center’s 2024 study in order to address the gaps in the two papers.

Gap

While existing research offers information about public opinions on the Electoral College, significant gaps remain in understanding teen perceptions, specifically, who will become the generation that takes over the nation and its institutions in the future. In the Pew Research Study mentioned previously, the population is exclusively adults, which fails to account for the perceptions of teenagers regarding this issue. This leaves a critical gap in understanding how high school students (teenagers), who are on the verge of entering the

electorate, perceive electoral systems. Additionally, the study by Maffucci regarding perceptions of college students, while accounting for a younger demographic, still overlooks the opinions and perspectives of high school students, which is what this study will aim to address. Furthermore, no previous studies, including Maffucci, have explored educators' perspectives, which play a critical role in shaping students' civic knowledge and attitudes. The researcher aims to fill these gaps by examining the perceptions of high school students and incorporating teachers' opinions on youth civic engagement and electoral attitudes. By addressing this often dismissed group and including multiple perspectives, this study provides a more nuanced understanding of how high schoolers perceive the Electoral College, having major implications on whether this century-long system will continue into the next generation.

III. Methods

Introduction

This study employed a comparative mixed-method approach to explore youth perceptions of the Electoral College and their support for reform. The research aimed to address gaps in the existing literature by testing three hypotheses that examined youth political activity, ideology, and sources of knowledge about the Electoral College. Data collection included student survey responses and in-person interviews with high school government teachers. The surveys, distributed via Google Forms, were designed to align with prior research, including Maffucci, Karp et al., Syversten et al., and elements of the Pew Research Center' study.

Hypothesis

To guide the research after identifying the gaps in the field of inquiry, three hypotheses were utilized:

1. Politically inactive youth will have less knowledge about the Electoral College compared to politically active youth.
2. Political ideologies will affect support for Electoral College reform.
3. Youth are more likely to gain their knowledge about the Electoral College from within the classroom than from outside influences.

The first hypothesis aligns with Maffucci, who discovered that politically active college students favored electoral reform or abolition rather than continuing with the current Electoral College system. The second hypothesis aligns with both Maffucci and Karp et al., who found that conservatives “are 10 percentage points less likely to favor Electoral Reform” than liberals (Karp et al. 776). The third hypothesis was based on Syversten et al., who suggested that schools play an important role in educating and promoting youth civic engagement.

To test these hypotheses, a comparative mixed-method study was conducted by collecting student responses via a Google Form (Appendix B), as well as in-person interviews with Calabasas High School Government educators (Appendices C, D, and E). Both the form and the survey included an Informed Consent Section (Appendix A).

Population and Sample Selection

Data was collected from CHS, a large, high-performing public high school in Southern California. CHS is a mixed-income school, with 23.4% of students being socioeconomically disadvantaged. 60.9% of the study's participants reported household incomes of over \$125,000 a year. It is also a co-ed school with students from 9th to 12th grade. With approximately 1,800 students, the school's ethnic makeup is as follows: 73.4% White (including Middle Eastern/North African origin), 11.3% Hispanic, 5.6% Asian, and 3.1% African American. The study's participants included 71.4% White, 8.3% Hispanic, 8.9% Asian, 6.0% African American, and 24.4% Middle Eastern/North African individuals (Note: Percentages exceed 100% due to participants identifying with multiple categories). 51.5% of respondents were male, and 48.5% were female. 22.9% of students surveyed were 15 years old, 38.2% were 16, 30.6% were 17, and 7.6% were 18. Interview participants consisted of three CHS government teachers, who consisted of 75% of the government teachers at the high school. One government teacher teaches AP Government, while two teach CP Government.

Instruments

A combination of quantitative and qualitative instruments was utilized to collect data for this comparative mixed-method study. The primary data collection tools included an online survey distributed via Google Forms and semi-structured interviews. The survey consisted of four distinct sections: an informed consent section, demographic information, political awareness questions, and targeted questions about Electoral College perception. A variety of question formats were used, including open-ended qualitative questions, multiple-choice questions, and Likert scale questions. The survey questions were adapted from prior research, such as Maffucci, Diemer, Syversten et al., and the Pew Research Center, with additional items developed by the researcher to address specific gaps in the existing literature. For the Electoral College perception questions portion, the researcher used definitions from Encyclopedia Britannica to inform participants about the various reform options to the system. Semi-structured interviews were conducted with questions guided by the gaps identified in prior research. These interviews were recorded using the Voice Memos app on Apple devices, ensuring their accuracy.

Table 1: Student Survey Questions

Student Survey		
Survey Questions	Measurement Scale	Source
Demographic Questions		
What is your age?	15, 16, 17, 18	Self-defined
What is your gender?	Male, Female, Nonbinary	Self-defined
What ethnicity/race do you identify	White/Caucasian, Black (Of	Self-defined

with?	African descent), Latinx, Asian/Pacific Islander, Middle Eastern/ North African	
What is your household income bracket?	Under \$25,000, \$25,000 - \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 - \$149,999, \$150,000 and above	Self-defined
Political Awareness Information		
Which political ideology do you lean toward? (You may scroll right)	Far Left/Progressive, Establishment Liberals, Democratic Mainstayers, Outsider Left, Stressed Sideliners, Ambivalent Right, Populist Right, Committed Conservative, Faith and Flag Conservatives	Pew Research Center
Would you consider yourself politically active?	Likert Scale: 1 (Not Active At All), 2, 3, 4, 5 (Heavily influenced)	Maffucci (2018)
If so, how? Check all that apply	Checkboxes: Voting, Protesting, Grassroots, Boycott, Join an Interest Group, Join a Party, Work on a Campaign, Following the news, Debate Club/ Model UN, Social Movement, Social Media	Self-defined
What motivates you to be civically engaged/politically active?	Short Answer	Syvertsen et al. (2011)
Electoral College Perceptions		
The information I know about the Electoral College comes from school	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Diemer (2012)
I feel confident in my knowledge about the Electoral College	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Maffucci (2018)
The information I know about the Electoral College comes from outside of school	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Diemer (2012)
Do you believe American Presidential elections are just and	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Maffucci (2018)

fair?		
Which political party do you think benefits most from the Electoral College?	Democrats, Republicans, Independents, I don't know	Self-defined
The Electoral College is an effective way to elect the President and Vice President of the United States	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Pew Research Center
I support states adopting the "Proportional Allocation of Electoral Votes" system as an alternative for electing the President through the Electoral College	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Self-defined
Do you support states adopting the "Congressional District Method" system as an alternative for electing the President through the Electoral College?	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Self-defined
Do you support adopting the "popular vote" system as an alternative for electing the President through the Electoral College?	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Pew Research Center
Do you support states adopting the "National Popular Vote Interstate Compact" system as an alternative for electing the President through the Electoral College?	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Self-defined
Do you support abolishing the Electoral College?	Likert Scale: 1 (Strongly Disagree), 2, 3, 4, 5 (Strongly Agree)	Pew Research Center
Rank your preferences regarding the different systems to elect the President of the US?	Multiple Choice Grid: 1. Electoral College 2. Proportional Allocation of Electoral Votes 3. Popular Vote 4. National Popular Vote Interstate Compact 5. Congressional District Method	Maffucci (2018)

Table 2: Teacher Interview Questions

Semi-Structured Interview Topics + Basis Questions
CHS Government Teachers
Electoral College Perception: Being a government teacher, how do you personally perceive the Electoral College? Did you think about it differently before being educated about it?
Influences on Youth Civic Engagement and Perception: Do you believe that in-school or outside-school factors influence youth civic engagement more?
Electoral College Perception: Do you believe in-school or outside-school factors influence youth perception of the Electoral College most? What are students' overall opinions about the Electoral College when it is discussed in class?
Teaching and Curriculum: In government classes, do you discuss different reform options to the Electoral College? Do students' opinions change before they learn about these options?
Teaching and Curriculum: How would you say most students learn about political processes, such as the Electoral College? What is a way to get more students more interested in these issues?

Implementation

To ensure a statistically relevant sample, a stratified random sampling method was the most effective way to target social science classrooms at CHS; closely aligning with Maffucci. Social science courses were also selected because they are a four-year requirement at the high school, ensuring all students are enrolled in at least one course within this discipline. The researcher contacted all CHS social science teachers via email to share the Google Form survey. They also contacted all CHS government teachers, with three out of four government teachers choosing to participate in an interview. To achieve consistency and eliminate any confounding variables, the researcher met all teachers during their lunch breaks. To ensure student participation, social science teachers posted the survey link on their respective Google Classrooms, enabling students to complete the survey independently using their own devices. Data collection spanned a two-week period and yielded 179 responses. Prior to completing the survey, all participants, and their parents, if under 18, were required to review and sign an informed consent form to ensure ethical compliance.

Data Analysis System

To construct histograms, data was analyzed using Google Sheets. The histogram constructing tool of Google Sheets, a spreadsheet application that is part of Google's free online Google Docs Editors suite, was used to analyze the data. A free, online website titled *Social Science Calculator* was utilized to generate both P-values and T-values for this study. High school students' overall perspectives were summarized via these analyses. These statistical

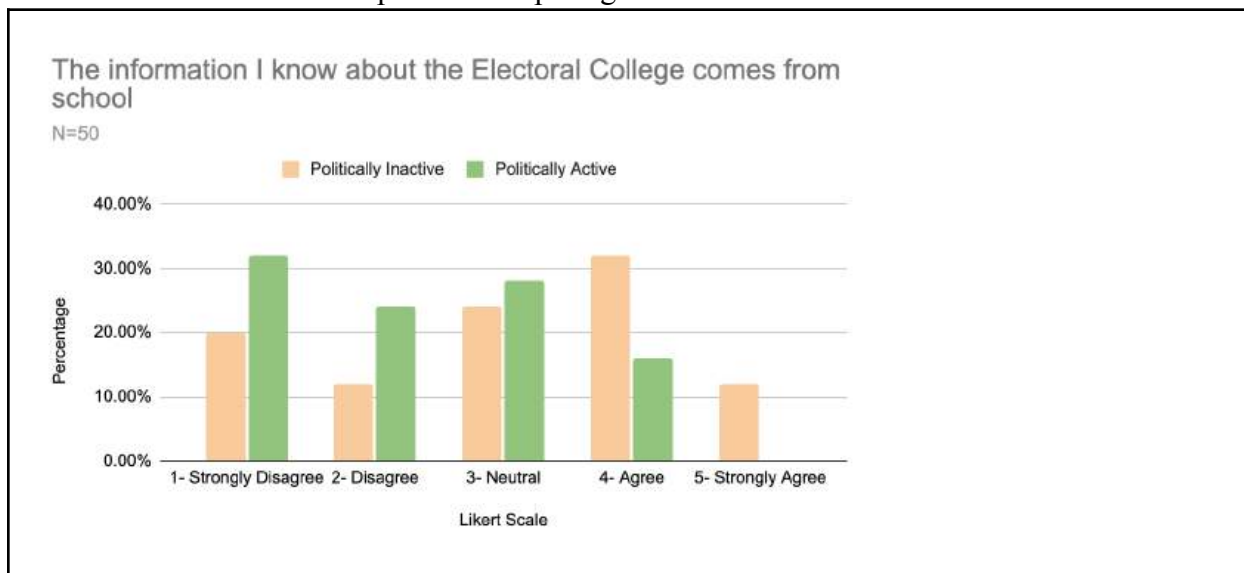
methods helped determine whether observed differences between groups were significant. For the interviews, the free website, Cockatoo.com, was utilized for transcription. The transcriptions were later placed in an online Word Cloud to highlight words or phrases that appeared several times in the interviews.

IV. Findings and Analysis

Subpopulations

The research highlights two key subpopulations: liberal-leaning and conservative-leaning youth, as well as politically active and politically inactive youth. These subpopulations are essential for understanding how ideological beliefs and political engagement shape opinions on the Electoral College. Political ideology plays a significant role in shaping perspectives, as shown in the 2024 Pew Research Center study and Karp et al. Additionally, political engagement can have major implications on youth perceptions of the Electoral College, which Maffucci documented in a previous study. By focusing on these two factors, the research aims to show how both ideology and engagement shape youth attitudes toward the Electoral College.

Table #3. Likert Scale Responses Comparing Political Active and Political Inactive Students



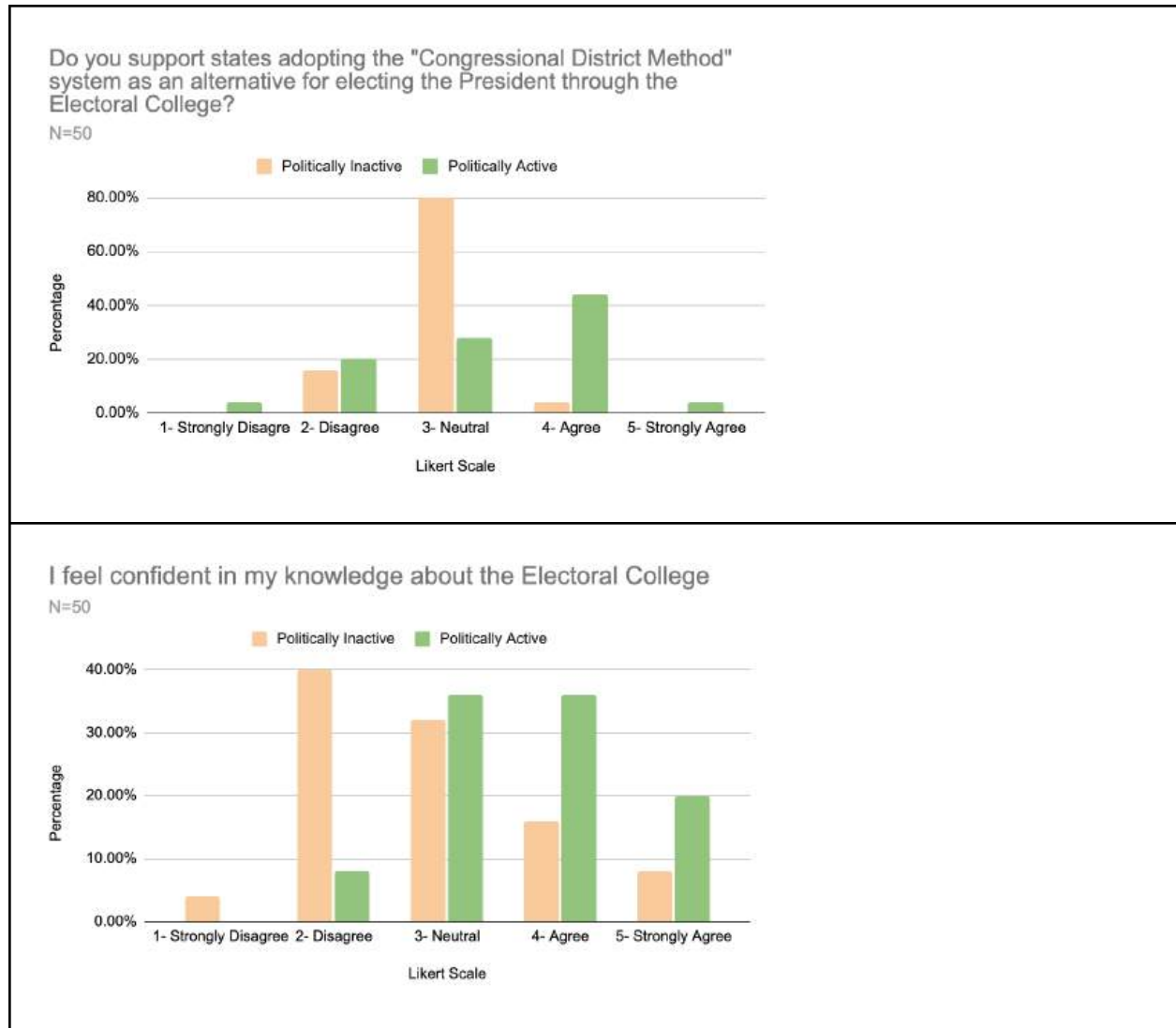


Table #4. T-Tests for Political Activity Questions above, *The result is significant at $p < .05$.

Questions	P value	T values	Average Inactive	Average Active
The information I know about the Electoral College comes from school	0.00844*	-2.567802	3.04	2.28
Do you support states adopting the "Congressional District Method" system as an alternative for electing the President through the Electoral College?	.03542*	1.890378	2.88	3.24
I feel confident in my knowledge about the Electoral College	.00669*	2.670595	2.84	3.68

Hypothesis #1: Knowledge between Politically Active and Inactive Youth

The data from the research shows that politically inactive youth are significantly more likely to rely on education from schools for information about the Electoral College. When asked, “The Information I know about the Electoral College comes from school,” politically inactive youth answered with a mean Likert Scale score of 3.04 compared to 2.28 for politically active youth. The highly significant p-value of 0.00844* indicates that this difference is statistically significant (Table 4). This suggests that politically inactive youth may primarily depend on in-school education for their understanding of the Electoral College, whereas politically active youth likely gain knowledge from a wider range of sources.

Beyond just where the two groups gain their knowledge of the Electoral College, politically active and inactive youth also show differences in their opinions on electoral reform. When asked about the Congressional District Method as an alternative for electing the president, politically active youth showed greater support, with a mean score of 3.24 compared to 2.88 for politically inactive youth. The statistically significant p-value of 0.03542* suggest that political activity does play a role in shaping opinions on electoral processes (Table 4). The Congressional District Method is among the most complex of the Electoral Reform alternatives, and therefore, politically active youth are likely to have more knowledge regarding it. Meanwhile, the lower level of support among politically inactive youth could indicate a lack of familiarity with the Congressional District Method, as shown by 80% of politically inactive youth choosing the “Neutral” option for this question, signifying their lack of knowledge about this system (Table 3).

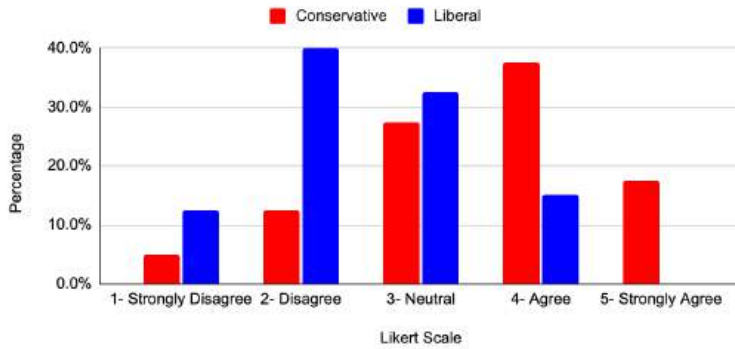
Confidence in knowledge of the Electoral College also varies significantly between politically active and inactive youth. Politically active youth reported significantly higher confidence levels, with a mean Likert Scale score of 3.68 compared to 2.84 for politically inactive youth. The significant p-value of 0.00669* confirms that this difference is statistically significant (Table 4). This suggests that political engagement not only increases exposure to information but also increases self-perceived understanding of electoral processes. Roughly 55% of politically active youth reported feeling confident about their knowledge of the Electoral College, showing that political activity directly correlates with knowledge of political systems (Table 3).

Overall, the research strongly supports the hypothesis that politically inactive youth will have less knowledge about the Electoral College compared to politically active youth.

Table #5. Perceptions Among Youth of Differing Political Ideologies

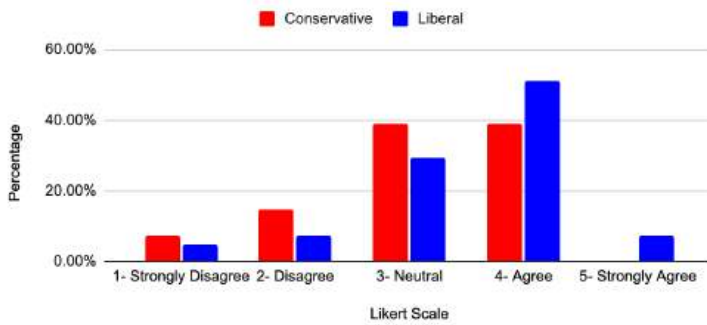
The Electoral College is an Effective way to Elect the President and Vice President of the United States

N=82



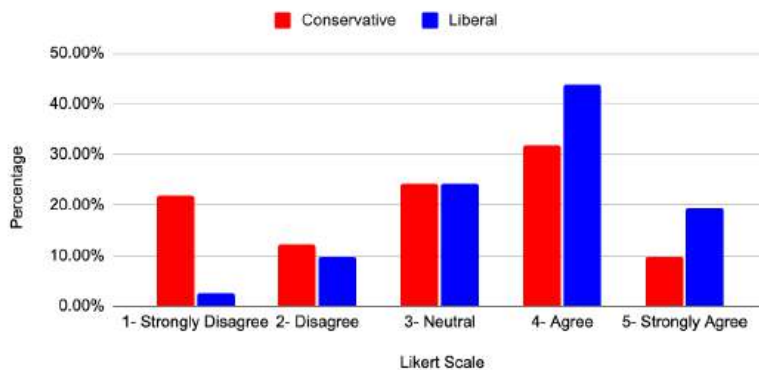
I support states adopting the "Proportional Allocation of Electoral Votes" system as an alternative for electing the President through the Electoral College

N=82



Do you support adopting the "popular vote" system as an alternative for electing the President through the Electoral College?

N=82



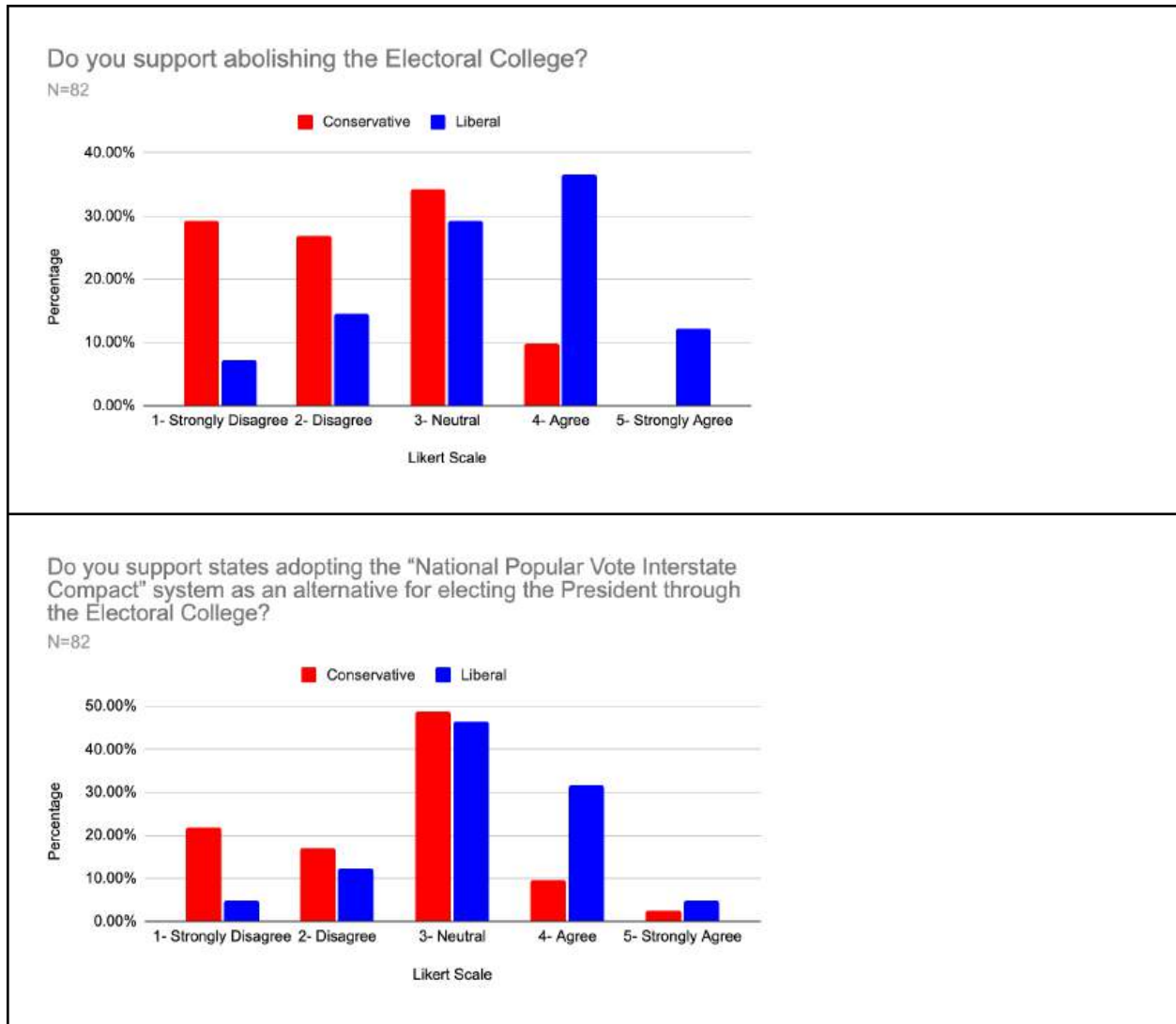


Table #6 T-Tests for the Political Ideology Questions Above, *The result is significant at $p < .05$.

Questions	P value	T value	Average Conservative	Average Liberal
The Electoral College is an effective way to elect the President and Vice President of the United States	0.00001*	-4.81206	3.56097561	2.512195122
I support states adopting the "Proportional Allocation of Electoral Votes" system as an alternative for electing the President through the Electoral College	0.02201*	2.079501	3.097560976	3.487804878

Do you support adopting the “popular vote” system as an alternative for electing the President through the Electoral College?	.00182*	3.088519	2.951219512	3.682926829
Do you support abolishing the Electoral College?	.00006*	4.286822	2.243902439	3.317073171
Do you support states adopting the “National Popular Vote Interstate Compact” system as an alternative for electing the President through the Electoral College?	.00175*	3.354102	2.536585366	3.195121951

Hypothesis #2: Electoral College Perceptions Among Youth of Differing Political Ideologies

The research strongly suggests that political ideology plays a significant role in shaping opinions on the Electoral College and potential reforms. Conservatives are more likely to agree that the Electoral College is an effective way to elect the President, with a mean score of 3.56 compared to 2.51 for liberals. The highly significant p-value of 0.00001* indicates that this difference is highly statistically significant. This finding aligns with Karp et al., who found that conservatives tend to favor maintaining traditional systems that they believe uphold stability and federalism. Liberals, on the other hand, are more skeptical of the Electoral College, likely due to the fact that many times, its outcomes do not align with the national popular vote (Karp et al.).

When examining support for the Proportional Allocation of Electoral Votes, a reform that would allocate electoral votes based on the proportion of votes each candidate receives in a state, liberals showed slightly greater support than conservatives, with mean scores of 3.49 and 3.10, respectively. The p-value of 0.02201* confirms that this difference is statistically significant (Table 6). This suggests that liberals are generally more open to reforms that move toward a more representative system. Conservatives, who are less supportive, continue to oppose Electoral reform alternatives.

The strongest ideological divide appears in responses related to abolishing the Electoral College and adopting a national popular vote. Liberals showed significantly greater support for the popular vote system, with an average score of 3.68 compared to 2.95 for conservatives, and the difference was statistically significant with a p-value of 0.00182* (Table 6). Similarly, when asked if they supported abolishing the Electoral College entirely, liberals had a significantly higher mean score of 3.32 compared to conservatives' 2.24, with a p-value of 0.00006* (Table 6). These results indicate a clear preference among liberals for moving away from the Electoral College, while conservatives remain more resistant to this change.

A similar pattern emerges when analyzing support for the National Popular Vote Interstate Compact (NPVIC), which seeks to ensure that the candidate who wins the national

popular vote becomes president without requiring a constitutional amendment. Liberals were significantly more likely to support this reform, with an average score of 3.20 compared to conservatives' 2.54. The p-value of 0.00175* confirms that this difference is statistically significant (Table 6). This finding could have implications for the NPVIC system as a whole, which requires the majority of states in America to be signatories in order to operate, meaning that if conservatives show resistance, this system is unlikely to succeed.

Overall, these findings strongly support the second hypothesis that political ideology influences support for Electoral College reform. Across all measures, liberals demonstrated significantly greater support for changes to the current system, particularly those that move toward a national popular vote, while conservatives were more likely to favor the existing Electoral College system.

Figure #1: Word Cloud of Government Teacher Interviews

TOP WORDS ⬇		BIGRAMS ⬇		TRIGRAMS ⬇	
Word	Frequency	bigram [®]	Frequency	trigram [®]	Frequency
school	22	i think	16	the electoral college	12
think	21	electoral college	14	outside school factors	10
factors	18	the electoral	12	factors influence youth	7
like	18	outside school	11	do you believe	5
outside	14	school factors	11	school factors influence	5
electoral	14	factors influence	8	or outside school	4
college	14	and then	8	of the electoral	4
i	13	do you	7	about the electoral	4
influence	11	influence youth	7	you believe that	4
kids	9	you believe	6	influence youth perception	3
parents	9	about it	6	youth perception of	3
in-school	8	of the	5	so i think	3

Hypothesis #3: Origin of Youth Knowledge Regarding the Electoral College

As shown in this Word Cloud, government teachers were asked whether their students tend to learn about political processes such as the Electoral College from in-school sources as opposed to outside-of-school influences. While various questions were posed during the interviews, the primary focus of this research is on the question: “Do you believe in-school or outside-school factors influence youth perception of the Electoral College most?” The Word Cloud analysis provides a clear answer to this question, demonstrating that government teachers overwhelmingly identified outside-school factors as the dominant influence on students’ perceptions of the Electoral College. The data from the Word Cloud supports this conclusion. The word "outside" appeared 14 times, whereas "in-school" was mentioned only 8 times, showing how teachers believe that outside influences carry far more weight in a student's understanding of the Electoral College than in-school influences do (Figure 1).

Additionally, the trigram "outside school factors" was mentioned 10 times, and the bigram "outside school" appeared 11 times, further reinforcing the idea that students rely more on external influences when forming their understanding of the Electoral College. This contradicts Hypothesis #3, which posited that youth are more likely to gain their knowledge about the Electoral College from within the classroom than from outside influences. Instead, the findings strongly indicate that external sources, such as social media, news, and family/friend discussions, play a more significant role in shaping students' perceptions than their in-school education does.

Discussion

Overall, the findings from this study align with previous literature that proved a correlation with Electoral College perception and political ideology and activeness. With regard to the first hypothesis made, the results of this study support the conclusion that politically inactive youth are less likely to be informed about the Electoral College than politically active youth. These findings align with the sample population of Maffucci, who found that those who share the sentiment that politics is unimportant to them are less likely to know about the Electoral College and its workings. Maffucci established that 33.3% of inactive youth felt confident in their Electoral College knowledge, while this study concluded that 24% of inactive youth felt confident, showcasing that the two studies yielded similar results. The results of this study additionally confirmed the second hypothesis made, showing how political ideology shapes perception of the Electoral College among youth. Maffucci established similar results, underscoring the power of political ideology in influencing Electoral College perception. When assessing overall support for the Electoral College among University of Maine students, Maffucci found that 15.3% of those with liberal leaning ideologies supported it, while 68% of those with conservative-leaning ideologies supported the system. This study found strikingly similar results, with 15% of liberals supporting the Electoral College and 59% of conservatives supporting it, respectively. Lastly, the third hypothesis was not supported by the research, which found that according to interviews with government teachers, youth are more likely to gain their knowledge about the Electoral College from outside school factors as opposed to in-school education. This rejects research done by Syversten et al., which established that political processes such as the Electoral College were more likely to be taught to youth at school versus at home.

Limitations

Although many substantive findings were brought to the surface from this study, numerous limitations are still present. The study categorizes youth as politically active or inactive, but this distinction may not fully show the complexity of political engagement. Some participants may be politically aware but not formally involved in activism or organizations, while others may participate in minimal activities but identify as politically engaged. For future research, a more precise approach to measuring political activity could help ensure that this

portion of the project is more precise. Additionally, a limitation of the study was the decision to interview teachers rather than students when assessing whether youth receive information about the Electoral College from school or external sources, especially with a sample size of only three teachers. While teachers can provide valuable information about what is covered in their curriculum, their perspectives do not necessarily reflect students' actual knowledge retention or perceptions. Furthermore, future research should survey youth about even more factors that may affect youth perception of the Electoral College, such as family influence and social media engagement.

V. Conclusion

The purpose of this study was to investigate the factors that influence youth perceptions of the Electoral College. More specifically, the paper focused on youth political activity, ideological perceptions, and in-school vs. out-of-school effects on Electoral College reform. By surveying 179 students at Calabasas High School and interviewing three Government teachers at this same institution, this study was able to successfully look into whether these factors have weight on youth perception of the Electoral College. Prior literature regarding the public's perception of the Electoral College, such as Maffucci and the Pew Research Center's study titled "Majority of Americans Continue to Favor Moving away from Electoral College" addresses college students' and adults' overall opinions regarding reform and factors that influence reform, but fail to account for younger demographics—specifically those aged 14-18. This study does just that, by specifically focusing on the factors that influence high schoolers' perception of the Electoral College. Additionally, three hypotheses were utilized to guide the research, with two being proven correct and one being proven incorrect due to the data collected. Pertaining to Hypothesis #1, political activity among youth was shown to have a considerable influence on their knowledge of reform options and the Electoral College system as a whole, as shown in Tables 4 and 5. Similarly, Hypothesis #2 revealed that differing political ideology preferences among youth had a large effect on their support for reform, specifically, with conservative-leaning youth supporting the traditional use of the Electoral College and liberal-leaning youth supporting more progressive reform, as shown in Tables 6 and 7. Lastly, Hypothesis #3 assumed that youth would derive more information about the Electoral College from within school as opposed to outside of school. However, through interviews with several government teachers, this hypothesis was proven to be incorrect, as the educators reported many students receiving their knowledge about this process from family, friends, social media, etc., as seen in Figure 1.

As a result of the data collected from this study, it is clear that numerous factors, such as political activity and political-ideological preferences, influence the perception of systems such as the Electoral College. It is also evident that high schoolers' support for Electoral College reform is on the rise. Researchers may find the results useful for understanding how civic education shapes political knowledge, and could build upon this study by including even more factors that may alter perceptions of the Electoral College. Ultimately, this research emphasizes

the importance of having more informed and engaged youth voters. As the future leaders and decision-makers of the country, young people must be able to challenge and address a variety of electoral topics, such as the hotly contested Electoral College. Addressing the gaps in political education and engagement will not only strengthen democracy but also ensure that the next generation is prepared to evaluate and shape the systems that govern them right now.

Works Cited

- Aldrich, John, et al. "Sophisticated and Myopic? Citizen Preferences for Electoral College Reform." *Public Choice*, vol. 158, no. 3/4, 2014, pp. 541–558. JSTOR, <http://www.jstor.org/stable/24507613>
- Chappell, Bill. "The Electoral College Debate, Explained." NPR, 4 Nov. 2024, <https://www.npr.org/2024/11/04/nx-s1-5173568/electoral-college-explained>
- Clark, Walter. "The Electoral College and Presidential Suffrage." *University of Pennsylvania Law Review and American Law Register*, vol. 65, no. 8, 1917, pp. 737–747. JSTOR, <https://doi.org/10.2307/3314432>
- Diemer, Matthew A. "Fostering Marginalized Youths' Political Participation: Longitudinal Roles of Parental Political Socialization and Youth Sociopolitical Development." *American Journal of Community Psychology*, vol. 50, 2012, pp. 246–256.
- Fisher, Dana R. "Youth Political Participation: Bridging Activism and Electoral Politics." *Annual Review of Sociology*, vol. 38, 2012, pp. 119–137. <https://doi.org/10.1146/annurev-soc-071811-145439>
- Hayden, Grant M. "The False Promise of One Person, One Vote." *Michigan Law Review*, vol. 102, no. 2, 2003, pp. 213–268. University of Michigan Law School Scholarship Repository, <https://repository.law.umich.edu/mlr/vol102/iss2/2>
- Karp, Jeffrey A., and Caroline J. Tolbert. "Polls and Elections: Support for Nationalizing Presidential Elections." *Presidential Studies Quarterly*, vol. 40, no. 4, 2010, pp. 771–793. JSTOR, <http://www.jstor.org/stable/23044851>
- Maffucci, Maria. *The Electoral College: A System "for the People?"* Honors thesis, University of Maine, 2018. Digital Commons @ University of Maine, <https://digitalcommons.library.umaine.edu/honors/350>
- Pew Research Center. "Majority of Americans Continue to Favor Moving Away from Electoral College." Pew Research Center, 25 Sept. 2024, <https://www.pewresearch.org/short-reads/2024/09/25/majority-of-americans-continue-to-favor-moving-away-from-electoral-college/>
- Public Policy Institute of California. *The Political Views of Young Californians*. Mar. 2024, <https://www.ppic.org/publication/the-political-views-of-young-californians/>
- Rolat, Adam. Testimony on the National Popular Vote Plan for Council of the District of Columbia. Brennan Center for Justice at NYU School of Law, 19 May 2010, <https://www.brennancenter.org/our-work/research-reports/testimony-national-popular-vote-plan-council-district-columbia>
- Sances, Michael W. "Legitimate Questions: Public Perceptions of the Legitimacy of US Presidential Election Outcomes." *Research & Politics*, vol. 10, no. 4, 2023. <https://doi.org/10.1177/20531680231206987>
- Syvertsen, Amy K., et al. "Thirty Year Trends in U.S. Adolescents' Civic Engagement: A Story of Changing Participation and Educational Differences." *Journal of Research on*

Adolescence, vol. 21, no. 3, 2011, pp. 586–594.

<https://doi.org/10.1111/j.1532-7795.2010.00706.x>

Whitaker, L. Paige, and Thomas H. Neale. The Electoral College: An Overview and Analysis of Reform Proposals. Congressional Research Service, Library of Congress, Jan. 2001.

Calabasas Students Perceptions of Pretty Privilege By Ella Meyers

I. Abstract

This study analyzes how pretty privilege manifests itself into every life through Calabasas High School Students. The main question investigated how and why perceived attractiveness leads to certain treatment in different contexts. This pushes a gap between how most existing studies on attractive bias and pretty privilege are grounded in regions where physical appearance is not emphasized to the same extent as in affluent, image-driven areas, like Calabasas. Much of the current research has been conducted in the South or rural areas, where cultural values may prioritize tradition, community, or academics over appearance where physical appearance does not hold social weight. Using a qualitative inductive thematic analysis of semi-structured interview questions and the use of facial stimuli in scenario situations, the research identifies recurring themes/theories. These include “halo effect,” “matching hypothesis theory,” “sociometer theory,” “implicit bias theory,” and “face perception theory.” The results indicate that teens in an image conscious community like Calabasas consistently associate attractiveness with likeability, demonstrating how regional culture can intensify the influence of beauty standards on everyday judgments. While this study highlights the immediate perceptions and biases tied to physical appearance, it remains important to explore how these perceptions evolve.

II. Introduction

There is an underlying problem with the phenomenon of pretty privilege. Pretty privilege is when generally attractive people get more advantages and opportunities (Laguyo et al. 2024). Throughout society, there has been an obsession with presentation, whether it's the newest clothes, the best hair, or a perfect smile, trying to achieve a “picture perfect look.” This phenomenon of gaining privilege from appearance is increasingly problematic due to complications with ethical opportunities and the influence it has on today's culture. In high school, this unspoken advantage shapes social dynamics, affecting relationships and perceptions. Patzer (2006) explores this paradox in his book, explaining how physical attractiveness can both empower individuals while simultaneously externalizing societal inequalities, unrealistic beauty standards, and cultural biases. This can impact the development of meaningful relationships, as beauty can create biases that overshadow deeper connections, affecting social growth and reinforcing superficial standards.

Biases placed in adolescent years can carry into adulthood and are similar to habits in that they stick, shaping how people view and accept people. This connection with appearance and biases extends to different branches, including education itself. People's perceptions cannot be controlled, and automatic biases can be placed based on physical impressions, which can stem from the surrounding culture. In Adam Howard's (2020) paper, a researcher at Colby College in Maine, he expands the current understanding of privilege by investigating how advantages shape identity formation and how students implement their privilege in daily life. To explore this

relationship, Howard studies the perspective of affluent students from private schools. The students' insights show that they use specific ways of thinking and behaving to explain their advantages in life by creating distinctions between social classes, building solidarity within their class, and making sense of the disadvantages faced by others. The author claims that these ways of thinking and acting are not just skills or strategies that wealthy students have learned but instead are important parts of their identity and self-perception. A possible cause of this problem is the community or ideals in which one surrounds themselves, as individuals base perceptions and biases on the environmental nature. These expectations placed on students can either be positive or negative, as these advantages are not based on quality work but on the appearance of their faces. This brings the researcher to the focus question of: "What are Calabasas High School Students' Perceptions of Pretty Privilege?" Exploration of this is through a qualitative, phenomenological study, as it explores perspectives and first hand experiences, which allows for the full understanding and a personal account of the privilege through beauty.

III. Literature Review

Physical attractiveness has been a topic of fascination, with its impact and research extending into different branches like social interactions, career opportunities, and even mental and physical health. While it is deeply rooted in society, promoting advantages for some, it also poses ethical challenges for others.

Current Research and Perceptions

Attractiveness is perceived differently depending on the context and age groups, shaping social dynamics influencing perceptions of pretty privilege. Borch et al. (2011) leads a team of researchers at the Department of Sociology at the University of Alabama and explored the role of physical attractiveness and aggression in impacting high school students' popularity. The hypothesis explored was that both attractiveness and aggression influence popularity. The methods used were peer evaluations with high school students in grade 12. A class roster was presented, and asked the students questions like "who starts fights?" This led Borch and her colleagues to find that attractiveness correlated positively with both overall and perceived popularity, while aggression negatively correlated with genuine social acceptance, highlighting differences in how adolescents are liked versus socially visible. Another point of view has been discussed by a researcher from Middle Tennessee State University, Chance (2023) discovered that appearance significantly influences hiring decisions, with confidence also playing a role in these decisions. Participants made hiring decisions based on a photograph and resumé and rated the applicants' perceived efficacy and confidence. Although attractiveness positively correlates with hiring choices, confidence only partially explains this relationship, pointing out the nature of bias in recruitment, trying to bring awareness to equal selection practices. Both Chance (2023) and Borch et al. (2011) explore perceptions with the use of stimuli in different contexts, but both confirm biases' prominent role in different settings.

Societal Bias

These advantages are usually celebrated and used by many, but researchers have criticized and critiqued the concept of this phenomenon and the potential for reinforcing bad societal biases. In this paper, Yonce (2014), a researcher from Smith College in Massachusetts, investigates levels of physical attractiveness as a characteristic of privilege and oppression. She goes through the concept of privilege which is done through a theory of intersectionality, meaning the idea that different areas of privilege and oppression interact with one another in different ways. Studies such as Rhodes et al. (2005) and Reis et al. (1980) show that society tends to perceive older men as more attractive than older women, illustrating the interplay of different demographics.

Aligned Methodologies

Research on attractiveness and its social impact is mainly scaled through collecting qualitative data through questionnaires, interviews, and surveys. For example, Dickey-Bryant et al. (1986), from the University of Georgia, focused on the relationship between occupational success and physical attractiveness through questionnaires. The hypothesis of this paper suggests there will be a positive correlation between facial attractiveness and occupational success. The instruments used included 129 voluntary students enrolled in a psychology course and 75 black and white photographs of military white men. They had students rate these men on a scale of 1-5 on an IBM answer sheet and for each pair of photographs, participants were asked questions like “Which of these two do you think you would like more?” The results of this study indicate that physical attractiveness plays a significant role in determining occupational success. This is similar to Borch et al. (2011) as both use facial stimuli and similar “most likely” questions to gain data.

At the Catholic University of the Sacred Heart, researchers La Rocca et al. (2012) described the concept of consumer attractiveness and created a process to develop a measure to assess it. This paper picked four existing B2B firms in Switzerland and Italy, with a sample size of about 36 to 44 participants across these four groups. During the sessions, participants discussed aspects they consider when assessing customer attractiveness. This was an unstructured conversation, resulting in 118 items representing the concept of consumer attractiveness. The next part consisted of surveys and interviews to collect further data and the participants were asked to identify their top 5 customers based on turnover to minimize bias. The last instrument used was a 20-item customer attractiveness scale and the results found a positive correlation between customer attractiveness and supplier satisfaction, suggesting that higher customer attractiveness is associated with greater supplier satisfaction.

Kenrick et al. (1980) at ASU conducts a different approach where the team researched how exposure to attractive stimuli, such as attractive cartoon characters or in media, impacts judgments of average attractiveness. It explores whether seeing attractive people makes others appear less attractive in comparison, a phenomenon known as the “contrast effect.” Three studies were conducted to test this idea. In the first study, male dormitory residents who watched TV

shows with exceptionally attractive female characters rated an average-looking female as less attractive compared to a control group who did not watch the show. Two additional studies replicated the results in controlled settings. The research suggests that mass media exposure to extreme attractiveness can lower assessments of average attractiveness and dating desirability, assuming the impact would be even greater as television and movies have integrated their way into everyday life.

Gap

These foundational sources push the gap between how most existing studies on attractive bias and pretty privilege are grounded in regions where physical appearance is not emphasized to the same extent as in affluent, image-driven areas. Much of the current research has been conducted in the South or rural areas, where cultural values may prioritize tradition, community, or academics over appearance. In contrast, living in a city like Los Angeles and attending school in Calabasas introduces a unique culture and environment in which physical appearance holds social weight. In Calabasas specifically, the influence of celebrity culture and media visibility not only elevates beauty standards but also normalizes cosmetic procedures among teens, using it to maintain status. Here, beauty is not just a bonus, it is a form of status that can shape group dynamics and influence opportunities.

In these settings, social visibility is often dependent upon conforming to beauty standards. However, as Borch's (2014) study revealed, visibility does not always translate to likeability or acceptance. An individual may be in the spotlight due to their looks, yet still subject to genuine friends or objectification. This disconnect between being seen and being valued reveals a deeper complexity between pretty privilege.

This cultural contrast and pressure to maintain appearances led the researcher to explore how pretty privilege manifests itself in cities like Calabasas or even Los Angeles as a whole. This study investigates how students are perceived, evaluated, and treated based solely on their physical appearance, and whether attractiveness acts as a gateway to privilege or a barrier to fair treatment. By experimenting in an image-conscious environment, the study's main goal is to uncover how beauty-biased treatment may be intensified, normalized, or even resisted with everyday high school interaction.

IV. Methods

To guide the research on the influence of pretty privilege at Calabasas High School, three hypotheses were developed:

1. Attractiveness influences how people perceive others in certain situations
2. Respondents who perceive high levels of personal attractiveness will correlate with high rates of social advantages.
3. Respondents will consistently choose the more attractive individual in scenarios involving positive attributes.

The first hypothesis is broader, proposing that individuals gravitate toward pre-biased, whether negative or positive, assumptions when in certain situations. This hypothesis aims to look at how these biases present themselves in different social situations and whether they change or stay the same through different contexts.

The second hypothesis builds on the idea that individuals who are frequently praised for their appearance may have a greater awareness and benefit from pretty privilege. This is done through phenomenological interviews, seen in Table 2, with open-ended questions to go deeper into their experiences and understanding.

The third hypothesis shows the role of attractiveness in decision-making and quick judgments. By presenting the participants with scenario-based questions shown in Table 3, it is hypothesized that more attractive people will be favored. This can further prove assumptions that associate attractiveness with positive traits.

Population

The population of this sample consists of students from Calabasas High School, a large, high-performing public school located in a suburban setting. The school has approximately 1800 students across grades 9-12 and is co-educational. Calabasas is a high-income area, though 23.4% of the high school students are considered socioeconomically disadvantaged.

The school's demographic breakdown is predominantly White (73.4%), with Hispanic (11.3%), Asian (5.6%), African American (3.1%), students identifying with two or more races (5%), and others (16%) making up the remainder. This could contrast with other schools with a more diverse, either culturally or economically, population.

This specific research focuses on all four grade levels (9-12), offering a broad perspective of how attractiveness can influence perceptions and outcomes in the academic and social world. The inclusion of Calabasas High School's socioeconomic and demographic characteristics can allow for exploration of how these traits can intersect with patterns observed in other literature.

Instruments

Informed consent was first needed and sent out to participants' parents and themselves if under 18. Once signed, a series of semi-structured interview questions was created to gain an in-depth understanding of participants' experiences and perceptions regarding attractiveness and its influence. The interview questions explore themes like personal encounters and beliefs with appearance-based treatment. Interviewing teens from Calabasas allows the study to directly capture how deeply these standards shape societal perceptions. By combining personal interviews with photos, the research can further prove the hypothesis of physical appearance being weighted in judgment.

These scenario questions presented are designed to identify bias and decision-making processes influenced by physical appearance. Participants are presented with hypothetical situations and facial stimuli to try to make contextual choices.

The photographs were created with AI and were asked to create three photographs based on American standards of what a 10, 6, and 4/10 are like. This ensured a fair and standard creation process. These facial stimuli feature all neutral expressions, with each having unique characteristics to try and identify pre-made stereotypes. These situations and features create biases that they associate with in their head, giving context that they did not have previously in the interview portion.

Sample Selection

Students enrolled in the Associated Student Body (ASB) were picked to participate in this study. The class consists of 32 students, all dedicated to serving and representing the school. Since every student in the ASB class participates in this study, the selection process ensured that all members of the target group were equally likely to be included. ASB was chosen to due its competitive interview and election process, which pretty privilege and/or physical appearance or status could play a role in.

Implementation

After participants were selected, either they or their parents were provided with an informed consent form to review and sign before the interviews could proceed. Once the forms were completed, a series of interviews were conducted. Participants were reassured that the data collected would be used solely for research purposes and encouraged to answer as honestly as they could without judgment. The process began with 7 open-ended questions designed to gain their opinions and experiences in depth. After completing this portion, participants were asked to evaluate three AI-generated photographs, rating each on a scale of 1 (least attractive) to 10 (most attractive) based on attractiveness, and the experiment concluded with 9 vignettes (scenario-based questions) to gain contextual perceptions.

V. Findings

The next step in the methodology was to perform a thematic analysis of the data obtained through interviews. A thematic analysis is a technique used to identify and analyze prominent themes found in data (Braun et al, 2006). In this case, a thematic analysis allowed for the development of themes in the participants' responses, answering the research question. To find the themes, an inductive approach was used, meaning the themes identified came directly from what participants shared, rather than being pre-decided (Patton 1990). This choice was used to find patterns in perception rather than trying to prove an existing model, allowing new insights that might not appear in Southern studies.

Table 5: Definition of Themes	
Theme	Definition

Halo effect	“occurs when our positive impressions of people, brands, and products in one area lead us to have positive feelings in another area.” Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016).
Sociometer theory	“Self-esteem not only assesses and responds to the quality and quantity of an individual's actual relationships, but also monitors their eligibility for various potential relationships.” Bale, Christopher, and John Arche (2013)
Matching hypothesis Theory	claims that people are more likely to form a committed relationship with someone equally attractive Jia, Tao, et al. (2015)
Implicit bias theory	“unconscious attitudes, stereotypes, or preferences that influence our understanding, actions, and decisions toward individuals or groups Zhu, V., & White, D. (2024).
Face Perception Theory	“Certain facial qualities can create immediate impressions, affecting judgments and interactions.” Zebrowitz, L. A., & Montepare, J. M. (2008)

Table 6: Summary		
Interviewee	Summary	Theme
#1	This participant is a senior and a pep commissioner. Getting compliments provides a temporary boost in self-confidence, leading to confidence that affects how you feel in social situations. Self-perception is what truly matters, and confidence only enhances that. First impressions rely on appearance, which can impact opportunities short term. Personal presentation influences social interactions based on reactions, and attractive individuals could face stereotypes.	Face Perception
	First impressions and presentation influence their perceptions of confidence and success. Moderate attractiveness is preferred for friendships, work, and everyday interactions. Confidence also plays a role in determining perceived attractiveness. The final ratings: 8/10, 7/10, 5/10	Halo Effect
#2	This respondent is a senior and the ASB vice president. Compliments boost how they feel and can make them feel different. Confidence heavily impacts beauty and can impact how you view others. Being attractive makes others more likely to engage with one another and gives social advantages. An example the respondent brought up would be if they were merging in traffic, people usually let them pass because of looking presentable.	Implicit Bias
	Their scenario responses attributed the more positive qualities to the best-looking respondent, like success and confidence. People gravitate towards more people they like both mentally and physically. Participants chose the most attractive for the friendship-related questions. The final ratings of the images are: 10/10, 7/10, 4/10.	Matching Hypothesis
#3	This respondent is a junior and is the junior class historian. Compliments are enjoyable, but if they do not regularly receive them, it does not determine self-worth. They believe beauty affects confidence mainly due to receiving more compliments. Believes more attractive people gain more opportunities than others due to likability. A specific example: if they are wearing a specific outfit, more people will go up or compliment them.	Sociometer Theory
	They strongly agree that attractiveness forms relationships, but not professional connections or friendships. In the scenario questions, they did the same as participant 2, linking positive	Face perception

	qualities with attractive people. Their final ratings of the images are 6/10, 4/10, 3/10	
#4	This respondent is a sophomore and is the class vice president. Compliments raise their confidence, but if they do not receive them, it does not lower self-esteem. Beauty affects confidence rather than vice versa, as attractive people are more likely to be drawn to each other than others. This leads to making it easier to form relationships and connections, but does not view it as a factor in friendships. This participant noticed that if they dressed more formally, teachers are likely to view them differently than dressed in a more casual or lazy outfit. Lastly, they believe being attractive draws the spotlight towards them more, constantly worrying about other people's opinions and views.	Face perception
	In the scenario, they picked all the positive examples with the prettier images. Their final rankings of the images are: 9/10, 8/10, and 6/10	Halo Effect
#5	This participant is a junior and is the clubs and community commissioner. Compliments make them feel good about themselves, but if they do not receive them, it does not affect self-esteem. They believe confidence affects beauty, as when they are the most confident, they receive the most compliments and attention. They also explained that attractive people do have more advantages, citing an example of more people letting them cut to the front in lines. On the other hand, being attractive can make forming relationships and friendships harder because of intimidation.	Halo Effect
	Their scenario responses attributed the more likable qualities to the best-looking respondent and the more competent qualities to the moderate one. Visual cues like grooming and expressions influenced their final rankings of the images: 8/10, 6/10, and 5/10.	Face Perception
#6	This respondent is a senior and is the asb president. They already have a high self esteem, so compliments feel good but define nothing. They do not believe attractiveness gives advantages but rather ethnicity. Appearing confident makes others want to be around them. Being attractive helps form friendships as many people want to be friends with people they find physically appealing.	Sociometer Theory
	In the scenario, attractiveness was strongly linked to likeability, confidence, and approachability. However, success was aligned with the least attractive person, suggesting an exception to bias. Overall presentation influenced their final rankings of the images: 9/10, 5/10, 5/10	Implicit Bias
#7	This respondent is a senior and is the senior class vice president. Compliments make them feel better but do not lower self perception. They believe confidence and beauty are interlinked and influence each other. They also believe attractiveness creates advantages particularly in the entertainment world. Strongly agrees that attractiveness forms relationships even if it is not done consciously.	Halo effect
	In the scenarios, attractiveness was linked to confidence and social desirability and success was attributed to the least attractive individual. Glasses and specific facial features influenced rankings, reinforcing unconscious associations between intelligence and attractiveness. Their final rankings are: 9/10, 6/10, 7/10	Implicit Bias
	This respondent is a senior and is the campus entertainment commissioner. Compliments help with self esteem while also fixing insecurities. Believes beauty affects confidence because self perceptions shape how you interact with others. Attractive people gain better opportunities, as they are more likely to be noticed/chosen. A specific example is if they put more effort into putting different beauty products on, they feel less self conscious about it and more people will likely interact. This helps with making connections and relationships, since people tend to associate looking good with positive traits. However, being attractive can serve as a disadvantage, especially for women in the workforce.	Halo Effect

#8	The moderate girl was preferred for success and interaction and the least attractive girl was assumed to care less about their appearance, reinforcing stereotypes about grooming and presentation in the scenarios. This leading to the final ranking of: 9/10, 8/10, 6/10	Implicit Bias
#9	This participant is a junior and the publicity commissioner. Compliments uplift them but do not hurt their self esteem if not received. Beauty affects confidence, if you feel attractive, you become more confident. Explicitly acknowledges that pretty privilege is real and people can be treated differently. An example is they are not putting in effort for school, people assume it comes from tiredness and make comments about their appearance. People unconsciously make judgments and form friendships based on attractiveness. A drawback can be an attractive individual being reduced just down to their looks.	Halo Effect
	In the scenario, attractiveness was linked to confidence, likeability, and approachability. Moderately attractive individuals were chosen for competence based roles, reinforcing the idea of being attractive reduces themselves down. The least attractive individual was assumed to care less, leading to the final rankings of: 10/10, 8/10, 7/10	Face Perception
#10	This individual is a senior and a pep commissioner. Compliments boost self esteem, but lacking them creates a feeling of little validation. They believe confidence affects beauty, the more confident, the more attractive. Looks play a major role in hiring and professional opportunities, as well as friendship rings. People who look like each other stay together, mentioning the term “fake blondes.” Attractive people are seen as less intelligent, especially in movies, reinforcing that stereotype.	Sociometer theory
	In the scenario, the less attractive and moderate girls were chosen for all positive qualities, like friendships and confidence. Suggesting natural beauty over extreme attractiveness. Picked the ones closest to them. Leading to the final rankings of 4/10, 7/10, and 8/10.	Matching Hypothesis
#11	This participant is a junior and is the athletics commissioner. Compliments boost their mood but do not affect self perception. Beauty and confidence are interlinked. Attractiveness provides advantages in certain careers like modeling. A specific example of certain advantages include the way they dress, as they are treated better dressed up than down. Treatment is more about effort in presentation rather than attractiveness. But, physical attractiveness plays a huge role in relationships and people can bond over things that can increase appearances. It's more of a disadvantage for women to be attractive.	Face Perception
	Moderation is ideal for women in a workplace setting, but being attractive is associated with social approachability. When looking at the images, everything is noted in the presentation like hair, skin, etc.. Leading to the final ranking of: 9/10, 7/10, 6/10	Face Perception
#12	This participant is a freshman and the freshman class vice president. Compliments make them feel good and increase self esteem which connects with beauty affecting confidence. Without being attractive, confidence is difficult to achieve. Being attractive proves to have advantages, as people are naturally drawn to appealing people, especially first impressions. Different in clothing choices, nicer dressed they are deemed as a “better kid.”	Halo Effect
	Moderate attractiveness is preferred in friendships, work, and professional settings, while extreme attractiveness is linked to confidence. The final rankings are: 9/10, 7.5/10, 6/10	Face Perception Theory
	This respondent is a senior and is the wellness commissioner. Compliments are not needed for validation but are a nice gesture. The more confident a person is, the more attractive they appear. Attractiveness plays a role in business settings, especially in a male dominated workplace. Brands also tend to favorably pick attractive people as the face of their company. Friend groups	Halo Effect

#13	tend to tie in with people with similar scales of attractiveness.	
	Less attractive individuals are seen as more dependable and attractive people can be seen as self centered. Participants chose the pictures that favored their type and relatability leading to the ranking of: 7/10, 6/10, 9/10	Matching Hypothesis
#14	This individual is a senior and is the asb secretary and treasury. Compliments do not affect them as they are already confident in themselves. Confidence affects beauty, as a confident person appears to have their life together regardless of how they look. Continuously, they attribute more opportunities based on confidence than looks. Today's workforce prioritizes diversity rather than attractiveness. However, a drawback of attractiveness/confidence is attracting haters out of jealousy.	Sociometer
	In the scenario segment, while attractiveness influences first impressions, other factors such as confidence and work ethic play a bigger role in long-term success. Specifically in the group project example, the participant responded with the most attractive person as they would want to do all the work themselves, contributing to this popular stereotype. This leads to their final rankings of 9/10, 6/10, 7.2/10	Implicit Bias
#15	This participant is a senior and is a publicity commissioner. Compliments do not make or break her day but can make her feel happier about herself when received. Confidence allows individuals to overcome barriers and become more beautiful. Beauty places a significant role in career opportunities, as someone attractive may be chosen over someone less attractive. Their increased confidence helped build better relationships and made more people drawn to them. A major drawback to being attractive is constantly being compared	Sociometer
	Confidence and effort in appearance influence perceptions of attractiveness. They associate more makeup and a polished look with higher confidence and prefer friendships with those who embrace natural beauty and appear more relatable. This leads to the rankings of the images: 9/10, 6/10, 7.5/10	Implicit Bias
#16	The response is a junior and the junior class president. Compliments do not affect self perception as they value their own opinion more, internal validation over external approval. Beauty affects confidence, as when people feel attractive, they become more comfortable with themselves. Attractiveness provides advantages in specific career fields like modeling or receptionist. Does not believe or has not experienced any differences in treatment based on appearance. Attractiveness also impacts male to female relations. Disadvantages include certain stigmas like assumptions.	Halo Effect
	Moderate attractiveness is ideal for most roles, while extreme attractiveness is linked to confidence. They also associate less attractiveness with lower self-care, using hair and grooming as a factor in his decisions. Their final decisions were: 8/10, 6/10, 4/10	Face Perception
#17	This respondent is the future asb advisor. Compliments are appreciated but not essential. A person can be theoretically beautiful, but without confidence, they may not display attractive traits or feel comfortable. Attractiveness is an advantage in western society, an example being modeling. As they became more confident, they formed more friendships but it only being a factoring decision in relationships.	Halo Effect
	A put together person and the vibe of confidence influences the scenario results. Attractiveness is linked to success, and moderate attractiveness is aligned with friendship and work settings. The final rankings: 8/10, 7/10, 6/10	Face Perception Theory
#18	This respondent is a senior and is the diversity commissioner. Receiving compliments boosts perception but not receiving has a neutral effect. Compliments can come from a place of encouragement but can also reinforce societal beauty standards. Beauty comes from within and	Halo Effect

	without confidence a person lacks it. Being attractive gives advantages with the way the world operates, an example being social media and modeling. Noticeable change in treatment after appearance transformation, especially with relationships.	
	They connect with individuals who do not prioritize looks. The most attractive person does have social advantages, but is not referenced. The moderately attractive person with success and smarts. Final choices: 8/10, 8/10, 9/10	Matching Hypothesis
	The participant is a senior and a campus entertainment commissioner. For some people, compliments might affect self perception more than them. Beauty affects confidence, when people feel more beautiful, they feel more confident. Beauty provides advantages in fields like modeling but beauty alone cannot carry someone in life. Dressing nicely and putting in effort leads to better treatment and romantic partners can be drawn to this. Attractive people could face stereotypes.	Halo Effect
#19	Attractiveness is associated with confidence and social preference. Moderate attractiveness leads to reliability and competence. Less attractive individuals do not put effort. The final decision: 9/10, 8/10, 5/10	Face Perception
	This participant is a junior and is the clubs and community commissioner. Compliments make them feel good, but not receiving them does not change how they see himself. If society perceives someone as attractive, they tend to feel better about themselves and experience less judgment. Being attractive is useful in careers like modeling or customer service. Has not experienced different treatment but believes females experience it commonly. Attractiveness also helps form friendships and relationships. Expected to maintain their looks and judged if not.	Halo effect
#20	Attractiveness is linked to confidence and social desirability. Moderate is viewed for professionalism and reliability. Stereotypes about appearances influence perceptions, such as assuming the least attractive person cares the least about their looks. Final scores are: 8/10, 6/10, 5/10	Implicit Bias

VI. Discussion

Theme 1: Halo effect

All participants acknowledged, whether subconsciously or consciously, that attractiveness provides social and professional advantages. They also matched desirable and more positive traits with the more conventionally attractive photograph, aligning with the researcher's third hypothesis. The majority linked opportunities with attractiveness in jobs like modeling or service roles.

For example, this connection to the workforce is evident in interview 17, where a participant noticed that "In the modeling industry, most of the people that you see are traditionally in a Western sense attractive. So I think that definitely can open the doors." Based on frequency, modeling was mentioned approximately 25 times across all interviews and scenarios, indicating a recurring association between physical attractiveness and career opportunities, particularly in appearance-based industries. It demonstrates how societal biases can translate into actual advantages, more specifically how attractiveness is not just a personal trait but a professional asset.

In Interview 3, they highlighted that attractiveness influences hiring decisions even if they have equal qualifications. This was also said in Interview 20, "If someone's hiring people and they need like a nice person for their customer service role or somewhere where people are

coming into the business, they're gonna want to go with the person that looks more nice because a lot of perceptions are like people looking nice they might be nicer like as their personality comes” These perspectives reflect the common assumption of the halo effect, explaining why people may subconsciously favor attractive individuals in hiring, because having better looks means having better traits.

In Talamas et al. (2016) study, their main focus is to investigate the influence of the attractiveness halo effect on perceptions of academic performance and guess the accuracy and when controlling for the halo effect, the accuracy for judgments improved, suggesting that attractiveness can distort judgements. This directly relates to the findings of this particular study, where participants believed that attractiveness could sway hiring decisions and signal personality traits. Both studies expose how the halo effect extends beyond academics and into real-life interactions and decision making, where choices are ideally made based on skill. However, these decisions are often influenced by attractiveness, which does not correlate with ability or competence, showing how deeply rooted pretty privilege can be.

Theme 2: Sociometer

Compliments were frequently described as having a positive impact on participants' self-esteem, especially when tied to appearance. Several participants expressed that compliments helped them feel more confident and valued. However, the absence of compliments did not necessarily lower self-esteem unless others around them were receiving praise, triggering comparison and exclusion. Others linked compliments to personal experiences, such as weight loss or makeup, and noted how receiving more attention during these times elevated self-confidence. This aligns with the researcher's hypothesis.

For example, Interview 10 shared, “Getting compliments boosts my self-esteem. I don't think not getting compliments negatively impacts me unless I hear someone else get complimented, and sometimes I feel like I'm the one left out of that.” Self-esteem is related to social feedback, and these drops in self-esteem could be from comparison to others, influencing how participants view their own attractiveness. A decrease in confidence may also affect how others perceive them, reinforcing the cycle of pretty privilege through both internal (self-perception) and external (social responses) feedback.

Interview 18 pointed out “If they change the way they look and they get compliments, a lot of their hard work feels validated and they can feel more comfortable within themselves.” Compliments serve as social validation, confirming their appearance is more socially acceptable. When individuals receive positive feedback, their internal “sociometer” registers social acceptance, increasing comfort and confidence. This reflects how appearance-based validation can play a significant role in shaping self-image, further illustrating how attractiveness and social approval interact to reinforce pretty privilege.

This aligns with Chance, M. (2023), who revealed that physical attractiveness significantly influences hiring decisions, with confidence also playing a role in these decisions. In many professional settings, confidence is often interpreted as a signal of leadership and

reliability, which are valued attributes. Confidence is not always internal or innate, it can be externally reinforced through appearance-based validation, such as compliments or increased attention. Participants described receiving compliments made them feel more secure and how little changes made led to security and assurance. The combination of both Chance and this study presents a feedback loop, attractiveness leads to more compliments and validation, which increases confidence, which in turn boosts how others perceive their professionalism or capability. Pretty privilege not only operates on just visual bias, but on the confidence beauty can generate, going far beyond looks.

Theme 3: Matching Hypothesis

Attractiveness plays a notable role in how individuals form social and romantic relationships. Most participants believed that people tend to seek out partners of similar attractiveness, whether subconsciously or consciously. Several participants acknowledged feeling pressure to befriend or date someone who reflects their attractiveness level, suggesting that perceived social compatibility is often influenced by social appearance, directly confirming the researchers second hypothesis. Additionally, confidence was repeatedly mentioned as a quality that can override this expectation.

An example from interview 12 shows “The more appealing you look, the more a person may want to receive or look at you, and want to talk to you.” Physical appearance can increase social approachability, drawing others in and increasing the likelihood of interactions. It reinforces the idea that attractiveness not only influences romantic connections but also plays a role in initiating everyday social connections.

Interview 20 emphasized the broad impact of attractiveness across different types of relationships, “I think it helps form all three connections honestly, because I think that if you look nice to society, then especially in a high school setting, a lot of people are going to want to be friends with you if you're like the popular nice-looking person.” This response shows how societal beauty standards can shape social inclusion, particularly in environments like high school, where image and popularity are often closely tied. It supports the idea that people are more likely to form connections with those who match or reflect socially desirable traits.

Interview 11 observed, “I think especially relationships, because obviously in relationships, a lot of people do care. Physical looks do play a role in it. So I think it does help in that circumstance. And also in friendship, because in friendship, you can bond over stuff similar to that.” Participants recognized that, while not the other factors, attractiveness still meaningfully shapes who individuals feel drawn to in both romantic and platonic relationships. The idea of bonding over shared or similar appearance suggests people feel more comfortable or more drawn to those who reflect their level or style.

Rhodes et al. (2005) did a similar experiment regarding if attractiveness influences dating success. Their findings suggested that attractiveness plays a crucial role in partners and is subject to sexual selection, meaning that physical appeal directly increases an individual's chance of being chosen as a partner. It concluded that attractiveness independently affects mating

success, showing that appearance alone can significantly shape romantic outcomes. This connects to the current research, as participants frequently acknowledge how appearance influences not only dating but also friendships and social status.

Theme 4: Implicit Bias

Many participants demonstrated unconscious bias regarding attractiveness, intelligence, and effort. A common pattern emerged where participants consistently linked glasses with intelligence, perpetuating a stereotype that has become deeply embedded in societal perceptions. Additionally, when responding to questions related to competence, the majority of participants selected the moderately attractive photograph, suggesting that participants may view extreme attractiveness as less credible or serious in contexts requiring skill or effort. These patterns reveal how subtle visual cues can trigger assumptions about abilities.

Interview 1 captured this bias directly, stating, “I think there could be some disadvantages, just based on generic stereotypes, maybe they think the more attractive you are, the stupid you are.” This reflects the stereotype that high attractiveness may come at the expense of perceived intelligence, believing that participants subconsciously correlate extreme attractiveness with incompetence.

Interview 2 stated a similar concern, “If you're attractive, people can hold that against you in some aspect. They'll have a preconceived notion of you or a perception of you that's not you, just because you look a certain way.” Combined with interview 1, this suggests that attractive individuals may face a negative bias, where their appearance leads others to underestimate them, leading to mischaracterization. These insights reinforce the complexity of pretty privilege, showing that while attractiveness often offers advantages, it can also trigger stereotypes that limit how people are intellectually perceived.

In interview 10, further emphasized this connection between appearance and intelligence, “I think people can sometimes think that you're less smart. There's a stereotype of looking smart and a stereotype of being dumb, and I think being dumb, you look slightly prettier, and being smart, you look uglier.” Demonstrates the internalization of stereotypes that tie beauty to superficiality and intelligence to lower looks. This perception helps uncover why many participants selected the moderately attractive individual in scenarios related to competence.

This is similar to Hosoda et al. (2003) as the purpose of this meta-analysis is to explore the bias effect of physical attractiveness in job-related terms and found that attractive individuals performed better than less attractive people in job-related outcomes. This aligns with the broader theme in this study, where participants acknowledged that physical appearance can influence perceptions of competence, confidence, and professionalism. On the other hand, unlike the uniformly positive outcomes highlighted in Hosoda et al. 's findings, the interviews and scenarios tested in this current study showed a more complex reality. While attractiveness can boost job prospects, it may also trigger negative stereotypes, proposing that attractiveness-based bias can both elevate and limit individuals, depending on the context and the traits being judged.

Theme 5: Face Perception

Most participants noted and presented that presentability and grooming significantly influenced attractiveness and perceptions. Rather than focusing solely on natural features, participants emphasized that being well-groomed, through clean clothing, styled hair, skincare, or other intentional appearances, could elevate how attractive someone appears to be. This suggests that attractiveness is often seen as manipulable, shaped not just by genetics but also by effort. This theme reflects the broader psychology and social cognition around face perception, where visual cues are interpreted to make judgments about personality. It reinforces the idea that attractiveness is not purely physical but also constructed through presentation.

In interviews 6 and 8, participants pointed out that messy hair or tired expressions make the participant seem less reliable and less attractive. One participant shared, “You can tell by hairstyles, like if they care about presenting themselves well, if they’re confident or not.” This shows how people make quick judgments based on subtle facial cues and grooming, interpreting them as effort. These impressions are often formed in seconds, supporting the idea that face perception goes beyond just features, but instead presentability. When individuals lack this, they can unintentionally give off a lack of care, which can impact them in social situations.

Interview 1 captured the immediacy of appearance-based judgments, stating: “Your first opinion is based on someone's appearance.” Face perception plays an important role in initial impressions, often before a person speaks or interacts. Physical appearance as a shortcut for assessing someone's character or status. In the context of this study, that means participants may not only recognize this tendency in others but also participate in it. This quote is not just an observation, it captures the origin that fuels social advantages based on appearance before any interaction occurs.

This is similar to French et al. (2009) who explored the relationship between grooming, personality, and academic performance. The findings revealed that being considered attractive alone positively impacts females' GPA. However, when personality and grooming are also considered, the research finds that it has the most impact on males. These results show that presentability can shape real outcomes like success, and that different appearance-related traits are rewarded differently based on gender. This connects directly to the current study, where participants consistently noted that grooming, confidence, and effort in appearance positively influenced how others perceive them. Therefore, society does not just reward natural features; it also values how well individuals manage their appearance, often interpreting grooming as a reflection of responsibility. This reinforces pretty privilege not just as a static trait, but something that can be earned or lost based on presentation, adding another layer to this bias.

VII. Conclusion

This study aimed to explore the influence of attractiveness on social and professional advantages, commonly known as “pretty privilege.” The main question investigated how and why perceived attractiveness leads to certain treatment in different contexts. Further, to answer the researchers' question, they analyzed three hypotheses: 1. Attractiveness influences how

people perceive others in certain situations. 2. Respondents who perceive high levels of personal attractiveness will correlate with high rates of social advantages. 3. Respondents will consistently choose the more attractive individual in scenarios involving positive attributes. These were all correctly proven with an inductive thematic analysis through a qualitative phenomenological approach with interviews and scenarios. The 5 key themes demonstrated how attractiveness influenced perceptions and outcomes in social, academic, and professional contexts, by applying them to a specific, highly image-conscious environment. This study fills a notable gap by focusing on Calabasas, a community heavily influenced by celebrity culture, media visibility, and social comparison. In this setting, beauty standards are not just present; they are amplified. By examining how students from this region perceive and respond to physical attractiveness, the study provides a lens into how pretty privilege functions where image is a form of status. Teens in an image-conscious community like Calabasas consistently associate attractiveness with likeability, demonstrating how regional culture can intensify the influence of beauty standards on everyday judgments.

Limitations

Despite its contribution to existing literature, limitations can be found. Similar to Shackelford, T. K., & Larsen, R. J. (1999), the sample size was relatively small as it only included participants from a single class, therefore, it may not represent broad populations. Additionally, responses were based on hypothetical situations rather than a real-life interaction, which could impact the authenticity of the participants' responses. This study also relied on self-reported data, leading some respondents to pick more "socially acceptable" answers rather than their true opinions. This was also seen in Chance, M. (2023), as their limitations consisted of their research being self-report, which affected the data with errors, such as rushing or lying.

Implications

Future research should explore the long-term effects of attractiveness across different domains such as the workplace, academia, and social settings. While this study highlights the immediate perceptions and biases tied to physical appearance, it remains important to explore how these perceptions evolve. Longitudinal studies could examine whether attractive individuals experienced sustained advantages or disadvantages. To promote fairness and equity, institutions and companies should implement blind recruitment, where visual cues or stimuli are removed from the application process to ensure decisions are based on qualifications. For example, A Harvard Business Review found that when applicants' gender and appearance information was anonymized, women's acceptance rates for interviews increased significantly, suggesting that biases linked to these visual cues may disadvantage certain groups (Fath 2023). Expanding these findings on a societal level reinforces the need to implement these practices, especially in environments like Calabasas where beauty and image are heavily emphasized from a young age, encouraging a cultural shift toward valuing competence and character over outward appearance.

Works Cited

- Bale, C., & Archer, J. (2013). Self-perceived attractiveness, romantic desirability and self-esteem: a mating sociometer perspective. *Evolutionary psychology : an international journal of evolutionary approaches to psychology and behavior*, 11(1), 68–84.
- Borch, C., Hyde, A., & Cillessen, A. H. (2011). The role of attractiveness and aggression in high school popularity. *Social Psychology of Education*, 14, 23-39.
- Chance, M. (2023). *Pretty Privilege at Work: The Influence of Physical Attractiveness on Hiring Decisions, Mediated by Perceived Efficacy* (Doctoral dissertation, Middle Tennessee State University).
- Dickey-Bryant, L., Lautenschlager, G. J., Mendoza, J. L., & Abrahams, N. (1986). Facial attractiveness and its relation to occupational success. *Journal of Applied Psychology*, 71(1), 16.
- Fath, S. (2023). When blind hiring advances DEI — and when it doesn't. *Harvard Business Review*.
- French, M. T., Robins, P. K., Homer, J. F., & Tapsell, L. M. (2009). Effects of physical attractiveness, personality, and grooming on academic performance in high school. *Labour Economics*, 16(4), 373-382.
- Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel psychology*, 56(2), 431-462.
- Howard, A. (2010). Elite visions: Privileged perceptions of self and others. *Teachers College Record*, 112(8), 1971-1992
- Jia, T., Spivey, R. F., Szymanski, B., & Korniss, G. (2015). An Analysis of the Matching Hypothesis in Networks. *PloS one*, 10(6), e0129804.
- Kenrick, D. T., & Gutierrez, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, 38(1), 131.
- Laguyo, V. J. G., Alibuag, J. J. L., Custodio, S. B., & Tagure, D. J. B. (2024). Strikingly Attractive: Social Experiences of Male Senior High School Students with Pretty Privilege. *Animo Repository*.
- La Rocca, A., Caruana, A., & Snehota, I. (2012). Measuring customer attractiveness. *Industrial Marketing Management*, 41(8), 1241-1248.
- Patzer, G. L. (2006). *The power and paradox of physical attractiveness*. Universal-Publishers.
- Reis, H. T., Nezlek, J., & Wheeler, L. (1980). Physical attractiveness in social interaction. *Journal of personality and social psychology*, 38(4), 604.
- Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success?. *Evolution and human behavior*, 26(2), 186-201.
- Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior*, 20(1), 71-76.

- Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PloS one*, 11(2), e0148284.
- Umberson, D., & Hughes, M. (1987). The Impact of Physical Attractiveness on Achievement and Psychological Well-Being. *Social Psychology Quarterly*, 50(3), 227–236.
<https://doi.org/10.2307/2786823>
- Xu, S., Martinez, L., & Smith, N. A. (2020). The effects of attractiveness, gender and self-esteem in service jobs. *International Journal of Contemporary Hospitality Management*, 32(1), 249-266
- Yonce, K. P. (2014). Attractiveness privilege: the unearned advantages of physical attractiveness.
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and personality psychology compass*, 2(3), 1497.
- Zhu, V., & White, D. (2024). Pretty Privilege vs. Group Prejudice... Who Wins?.

PhantomRx: Alleviating Phantom Limb Pain Using Bionic Hand

By Ansh Saraf, Ti’Air Riggins

Abstract

Phantom limb pain (PLP) arises when the brain sends signals to an absent limb, often resulting in pain, tingling, or discomfort. While various approaches exist to manage PLP, there are limitations. This paper seeks to provide a more accessible and affordable solution, the PhantomRx, for PLP relief. By developing a low-cost, 3D-printed sEMG (surface electromyography) bionic hand, this project aims to offer visual feedback that simulates the movement of a missing limb, potentially alleviating PLP symptoms. The sEMG sensors detect muscle signals in the residual limb, helping PhantomRx to drive the corresponding movements of the bionic hand. This completes a visual feedback loop that mimics real-time hand movements which relaxes the neurons. Additionally, PhantomRx intends to synergistically use TENS (transcutaneous electrical nerve stimulation) as a complementary pathway for PLP relief. TENS focuses on stimulating the nerves to provide reverse feedback directly in accordance with desired movements. The engineering challenges involved designing and manufacturing a functional, low-cost bionic hand suitable for PLP relief. Key design considerations included selecting cost-effective yet reliable materials, determining appropriate motor types, optimizing design features for comfort and durability, and enabling gesture recognition to simulate natural hand movements. This holistic approach not only seeks to address the functional needs of amputees suffering from PLP, but also focuses on cost optimization, creating a potential solution for effective and accessible pain management.

1. Introduction

Recent research suggests over 10 million people worldwide live with amputations; of those 10 million, approximately 85% experience Phantom Limb Pain (PLP) (Understanding Phantom Limb Pain: Causes, Symptoms, and Effective Management);(Understanding Phantom Limb Pain: Causes, Symptoms, and Effective Management; Crawford). Specifically, for those with wrist amputations this pain often presents itself as a persistent sensation or discomfort, since the brain continues to perceive the missing limb as though it was still present (Ramachandran and Blakeslee). This issue has been ignored far too long, and it is imperative to make it a key area of focus in neural engineering (Talks).

This project (PhantomRx) aims to present a more accessible and affordable solution to relieve phantom limb pain by developing a low cost, 3D printed sEMG bionic hand (Velasco). The objective is to move the bionic hand when the patient intends to move the missing limb. The visual feedback from this movement tricks the brain into viewing it as the actual limb, hence possibly relieving phantom limb pain (“Mirror Therapy”). The sEMG (Surface Electromyography) sensors detect muscle signals on the residual limb, driving the movement of the bionic hand mimicking the desired hand movements (“EMG: Electromyography”)Additionally, the project also focuses on research into Transcutaneous

Electrical Nerve Stimulation (TENS) as an alternative/supplementary approach for PLP relief. TENS sends small electrical currents through the affected nerves stimulating them and embodying muscle movements providing the electrical feedback loop to the brain. This essential function communicates sensory feedback from the prosthetic to the brain, which has been proven to alleviate or lessen phantom pain experienced by amputees (Limakatso and Parker); (Ramachandran and Blakeslee)

There are different hypotheses on why people experience phantom limb pains. The dominant theory is related to cortical reorganization of the neurons in the somatosensory context (Schone et al.)(Ramachandran and Blakeslee). It states that the neurons dedicated to the lost limb are reorganized and respond to spurious signals from other neighboring neurons.

Another hypothesis suggests that the brain stores an internal map (Cortical Homunculus) or representation of the body. This representation still remains intact even after the limb is lost and this discordance and lack of sensory feedback causes sensation and pain (- YouTube). It implies that pain, in the absence of ongoing tissue damage, might be caused by incongruence between motor intention and proprioceptive feedback (Website). According to the research, when a limb is intact, motor commands to move a limb are usually damped by sensory error feedback, such as vision and proprioception. With a phantom limb, this damping effect is not present and the motor output may become amplified and experienced as painful (Talks); (Talks; Ramachandran and Blakeslee). Interestingly, some infants born without limbs also feel phantom sensations and pain. Such evidence lends credence to this theory.

A third hypothesis suggests that when a limb is lost, the damaged neurons try to repair themselves and regrow, but without the limb to grow into, they form a massive mass of neural tissue called neuroma ((Understanding Phantom Limb Pain: Causes, Symptoms, and Effective Management); (Current Understanding of Phantom Pain and Its Treatment). The neuroma may generate erratic signals which cause phantom sensations and pain, which reduces the quality of life for patients. As result, recent advancements in f-MRI (functional magnetic resonance imaging) have been used to address PLP. f-MRI uses brain mapping to access the activity of multiple areas of the brain, while patients are performing various tasks. It is observed that the areas associated with the missing limb still light up when amputees experience phantom sensations or pain (Limakatso and Parker).

Treatments currently in practice to help with phantom limb pain are Mirror Therapy and Visual Image Therapy (Website) (Moseley et al.). They both rely on tricking the brain into believing that the amputated limb is still intact. There are also experiments which rely on VR headsets to create a similar illusion (- YouTube). Mirror Therapy is currently the gold standard being the most widely used therapy to relieve PLP. In this, the patient places the unaffected limb into one side, and the stump into the other. The patient then looks into the mirror on the side with an unaffected limb and makes "mirror symmetric" movements, as a music conductor might, or as we do when we clap our hands. Because the subject is seeing the reflected image of the good hand moving, it appears as if the phantom limb is also moving. Through the use of this artificial visual feedback, it becomes possible for the patient to "move" the phantom limb, and to unclench

it from potentially painful positions (Ramachandran and Blakeslee). This approach exploits the brain's preference to prioritize visual feedback over somatosensory/proprioceptive feedback concerning limb position (Talks)(Crawford).

The goal of this work is to use these underlying feedback principles to address PLP by combining engineering with therapeutic techniques in a single, low-cost solution. A 3D-printed sEMG bionic hand will not only function as a prosthetic but also provide visual feedback, simulating real hand movements to address phantom limb sensations (Velasco) (Cao et al.). Unlike traditional mirror therapy, which has limited accessibility and variable effectiveness, this bionic hand delivers dynamic feedback directly tied to the user's own muscle signals, creating a more immersive and potentially more effective experience (Wang et al.).

This project adopts a two-pronged approach to treating Phantom Limb Pain by combining visual feedback **and** sensory feedback. Visual feedback—through mirror therapy, VR, or movement of a bionic hand—has been shown to reduce PLP by aligning motor intention with visual perception. Sensory feedback is provided via TENS, which stimulates residual nerves and has been found to reduce pain by reintroducing somatosensory input(Current Understanding of Phantom Pain and Its Treatment)(Website). Research also suggests that adding vibration or haptic cues may further enhance this effect (Wang et al.). By combining both methods, this prototype, PhantomRx offers a comprehensive, low-cost potential solution grounded to address phantom limb pain while encouraging patients to regain some level of independence.

2. Materials and Methods

The bionic hand was built using two SG90 (TowerPro) and three MG996R (TowerPro) servo motors, responsible for finger and wrist movements. An Arduino Uno, programmed via the Arduino IDE, controlled the system, and a 2S 1000 mAh LiPo battery powered the electronics. Jumper wires and a breadboard, as illustrated in **Figure 1**, were used during prototyping, while a soldering iron finalized electrical connections. The structure, as illustrated in **Figure 2**, including the wrist, palm, four fingers, and thumb—was designed in Fusion 360 and 3D printed using a Prusa MK3s+ with PLA and HIPS filament. Nylon fishing string acted as tendons for flexion, while 3mm elastic string enabled passive extension. CA glue helped in assembling mechanical parts. EMG sensors and electrodes allowed for muscle-based control, translating forearm signals into servo movements, highlighted in the following ways. Each finger was segmented into many parts to allow for natural bending and closely resemble the action of the human hand. Five separate servo motors, each tasked with controlling a finger for autonomous movement, were housed in the wrist module's small form factor. The sEMG sensors were placed on the wrist over what is considered to mimic the radial and ulnar nerves in order to record electrical activity from residual nerves. This allows for a scalable, low cost, flexible, long-lasting precise signal recognition that may fulfill the functional needs of patients.

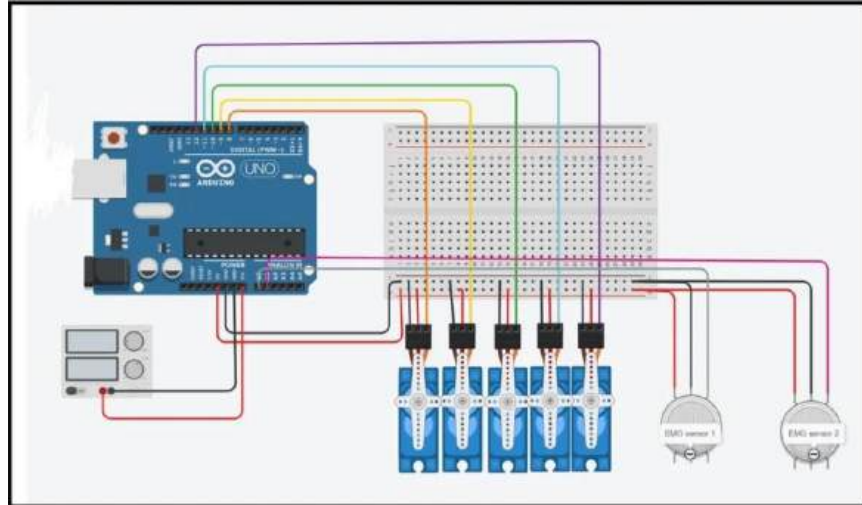
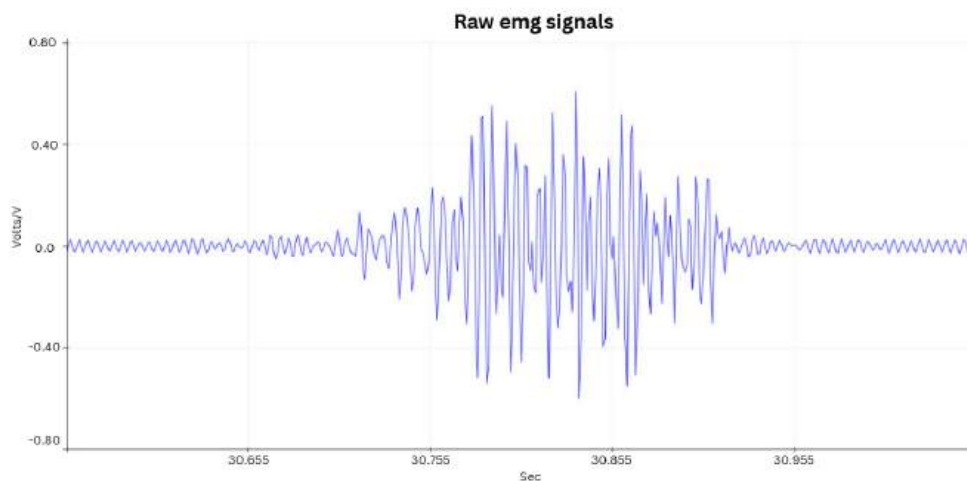


Fig 1. Circuit Diagram for bionic hand

The Arduino board was used to process the data after the sEMG sensors were positioned on the wrist and arranged inside the wrist module to reduce clutter and interference. Each finger movement was then calibrated and tested, with alignment and responsiveness checked to make sure the hand could mimic gestures in response to sEMG signals. In the signal acquisition step, electrical activity generated by muscle movements is captured by sEMG sensors placed on the wrist. Each sEMG sensor is equipped with three electrodes: a reference electrode, a positive input electrode, and a negative input electrode. The reference electrode stabilizes the signal by reducing noise, while the positive and negative electrodes capture the differential voltage created by muscle contractions. The sensors detect muscle activity along the radial and ulnar nerves, converting this into an electrical signal. This signal is transmitted through the sensor's data output pin, which is connected to an analog input pin on the Arduino processor. As a result, the Arduino receives the muscle activity data in its raw analog form, representing variations in muscle contraction intensity over time.



a

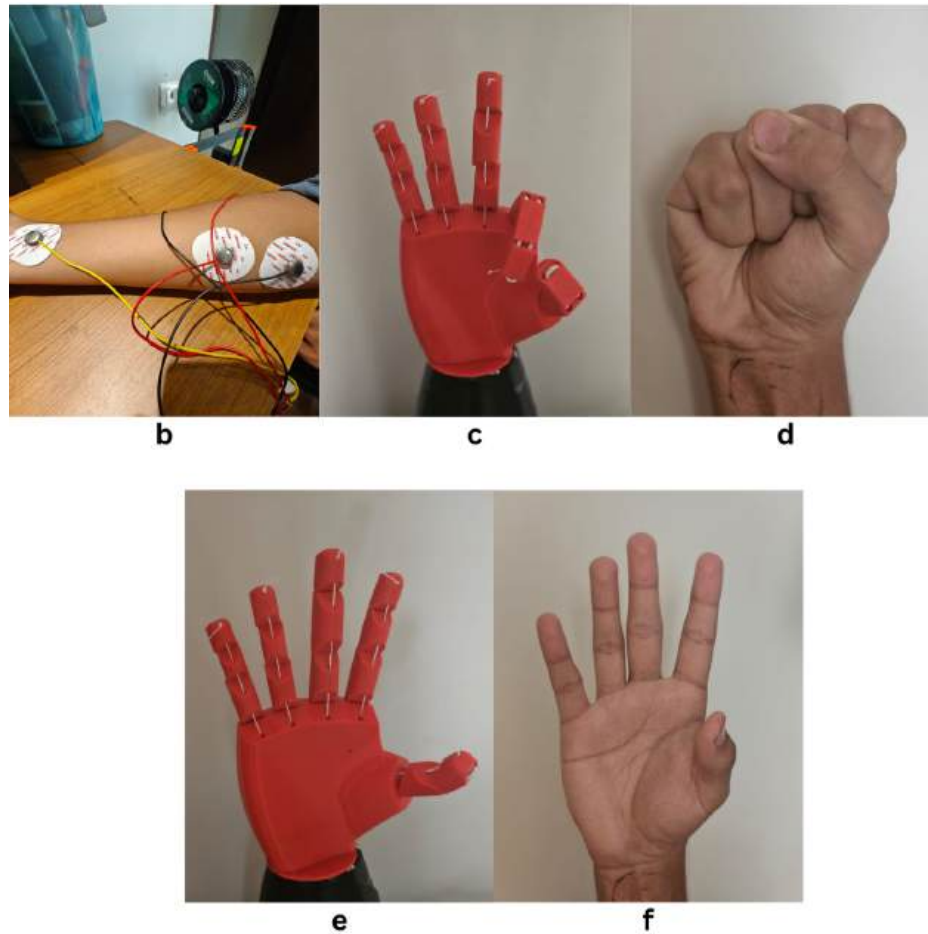


Fig. 2.

Fig 2. a. Raw emg signal (Mv vs Sec) **b.** Electrode placement on patient hand **c/d.** Wrist Clench The clenching gesture was recognized when there was increased activity in both the sensor on the radial nerve and the ulnar nerve. Once the wrist was clenched, the electrical signal increased from its resting value on both sensors but it increased by a lower degree than whilst only extending or flexing. **e/f.** Wrist extension The extensor gesture is recognized when there is high activity on the radial nerve. Once the wrist was extended there was minimal activity on the Ulnar nerve, however there was significantly high activity on the sensor placed on the radial nerve. Once an electrical signal from that sensor crossed a certain percentage threshold above the mean value (at rest) the extensor gesture was recorded.

In the signal processing step, the raw analog data received from the sEMG sensors represents fluctuations in muscle activity but is initially too erratic for reliable gesture recognition. To interpret these signals accurately, the data is converted into a continuous waveform, which smooths out rapid variations and highlights consistent muscle activity patterns. A digital low-pass filter is applied to achieve this smoothing effect, allowing only the lower frequency components of the signal—representative of muscle contraction patterns—to pass through while reducing high-frequency noise.

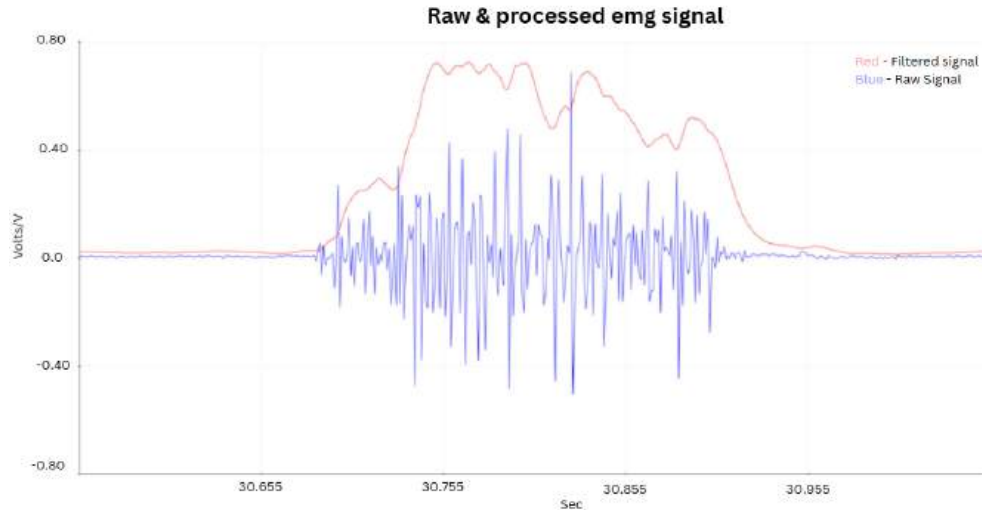


Fig. 3. Processed & Raw emg signal

The low-pass filter operates with a cutoff frequency set to capture muscle activity effectively while ignoring rapid fluctuations that do not contribute to gesture recognition. This filter performs an exponential moving average, smoothing the raw data and forming a continuous waveform. This processed data can now be used for gesture recognition by comparing the amplitude and patterns of the wave, enabling the system to identify specific gestures such as extensor, flexor, and clenching movements accurately.

3. Results and Discussion

Gesture recognition requires the usage of all 5 servo motors embedded for each finger for innervation. When the flexor gesture is recognized, the servos rotate to 180 degrees to close all fingers and form a fist. When the extensor gesture is recognized, all motors were released and returned to 0 degrees to open the hand completely. Once the clenching gesture is recognized, only the index finger motor and thumb motor are activated to 180 degrees to simulate a pinching movement.

To evaluate the performance of the bionic hand, three participants from different age groups were selected to ensure diverse neuromuscular responses, summarized in **Table 1**. The first participant was a 16-year-old male, the second a 38-year-old female, and the third a 64-year-old male. Each individual was instructed to perform three distinct gestures—Flex, Extensor, and Clench—20 times each, resulting in 60 trials per person. During each trial, surface EMG signals were recorded and processed to classify the intended gesture, which was then actuated by the prosthetic hand. The system's performance was assessed by comparing the intended gestures with the actual output, allowing the calculation of true positives, false positives, false negatives, and true negatives for each user.

Participant A

Gesture	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)	Accuracy (%)
Flex	18	1	1	40	96.70%
Extensor	17	2	1	40	95.00%
Clench	19	0	1	40	98.30%
Total	54	3	3	120	96.00%

Participant B

Gesture	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)	Accuracy (%)
Flex	17	2	1	40	95.00%
Extensor	16	3	1	40	93.30%
Clench	18	1	1	40	96.70%
Total	51	6	3	120	93.50%

Participant C

Gesture	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)	Accuracy (%)
Flex	19	1	0	40	98.30%
Extensor	18	1	1	40	96.70%
Clench	17	2	1	40	95.00%
Total	54	4	2	120	96.50%

Table 1. Participants A,B & C results for gesture recognition accuracy

4. Conclusions

This experiment revealed that the sEMG sensors and the 3D printed hand works synergistically, successfully recognizing and imitating the 3 gestures (flexor, extensor and clenching) with high accuracy, as demonstrated in **Table 1**. The durable joint systems also displayed accurate spatiotemporal dexterity, closely resembling natural hand movements. The material choice of nylon thread paired with elastic string also proved to be the correct one, as the elastic generated least resistance while providing greatest restoring force while the nylon survived repeated use without deformation or wear.

Over the long term, this translates to a potential prosthetic solution that is not only responsive and precise but also reliable under daily, real-world use. Patients using the hand over extended periods would benefit from the low-maintenance nature of these materials, which reduces the likelihood of wear-induced failure or the need for frequent recalibration. The consistent mechanical integrity ensures that repeated motion-such as daily grasping or flexion cycles-would not degrade performance or compromise safety.

Nylon, in particular, is an excellent choice for such applications. It offers high tensile strength, flexibility, and abrasion resistance, making it ideal for simulating artificial tendons or ligaments that undergo continuous motion and load. Additionally, nylon is lightweight and partially recyclable, which supports both patient comfort and environmental sustainability (“The Selective Dissolution-Precipitation Recycling: A Review on This Alternative Approach for the Recovery of Nylon-Based Waste Materials”). Its fatigue resistance has been validated in biomedical contexts, especially in external wearable systems, where it withstands repeated strain without loss of shape or function (Lynch-Aird and Woodhouse). Nylon is also known for being skin-compatible and non-irritating, an important consideration for prosthetics that stay in contact with the user for hours at a time (Lynch-Aird and Woodhouse; Shakiba et al.).

Polylactic acid (PLA) is a widely used biodegradable thermoplastic known for its excellent biocompatibility. Derived from renewable resources like corn starch or sugarcane, PLA naturally degrades into lactic acid, a compound safely metabolized by the human body. This makes it especially suitable for temporary biomedical applications such as sutures, implants, and scaffolds for tissue engineering. PLA has been shown to support cell adhesion and proliferation, making it a strong candidate for use in regenerative medicine and skin-contact prosthetic components (da Silva et al.). Moreover, its low toxicity and minimal inflammatory response ensure safe integration with biological tissues during short- to medium-term use.

The gesture recognition code achieved nearly a 95% success rate with minimal false positives, as the electrodes were easy to place and adjust. In addition, research into TENS shows great potential in replacing the conventional mirror box therapy as it targets the root cause of phantom limb pain. It could prove to be a potential long-term passive solution without the need for specialized training or personnel. Coupled with visual feedback from the bionic hand, the electrical stimulation has great potential to be an effective solution for alleviating PLP.

These findings underscore the potential for PhantomRx to serve as both a functional prosthetic and a therapeutic device for mitigating phantom limb pain. Future studies should explore the system’s long-term usability with actual amputee patients, quantify the extent of PLP relief over extended durations, and integrate a more refined sensory feedback (such as pressure or temperature) to deepen the brain-body feedback loop. Additionally, enhancing the visual realism of the prosthetic and exploring wireless or mobile-connected versions may improve user acceptance and accessibility, making PhantomRx a viable clinical tool in low-resource settings.

Works Cited

- Cao, Junyi, et al. "Hand Gestures Recognition Based on One-Channel Surface EMG Signal." *Journal of Software Engineering and Applications*, vol. 12, no. 9, Sept. 2019, pp. 383–92, <https://doi.org/10.4236/jsea.2019.129023>.
- Crawford, Cassandra S. *Phantom Limb: Amputation, Embodiment, and Prosthetic Technology*. NYU Press, 2014, <https://play.google.com/store/books/details?id=Z5xgAgAAQBAJ>.
- Current Understanding of Phantom Pain and Its Treatment. <https://www.painphysicianjournal.com/current/pdf?article=NzU1MA%3D%3D&journal=147>. Accessed 7 Aug. 2025.
- da Silva, Dana, et al. "Biocompatibility, Biodegradation and Excretion of Polylactic Acid (PLA) in Medical Implants and Theranostic Systems." *Chemical Engineering Journal* (Lausanne, Switzerland : 1996), vol. 340, Jan. 2018, p. 9, <https://doi.org/10.1016/j.cej.2018.01.010>.
- "EMG: Electromyography." BIOPAC Systems, Inc., 18 June 2015, <https://www.biopac.com/application/emg-electromyography/>.
- Limakatso, Katleho, and Romy Parker. "Treatment Recommendations for Phantom Limb Pain in People with Amputations: An Expert Consensus Delphi Study." *Pm & R*, vol. 13, no. 11, Feb. 2021, p. 1216, <https://doi.org/10.1002/pmrj.12556>.
- Lynch-Aird, Nicolas, and Jim Woodhouse. "Mechanical Properties of Nylon Harp Strings." *Materials*, vol. 10, no. 5, May 2017, p. 497, <https://doi.org/10.3390/ma10050497>.
- "Mirror Therapy." *Physiopedia*, https://www.physio-pedia.com/Mirror_Therapy. Accessed 7 Aug. 2025.
- Moseley, L. G., et al. "Is Mirror Therapy All It Is Cracked up to Be? Current Evidence and Future Directions." *Pain*, vol. 138, no. 1, Aug. 2008, <https://doi.org/10.1016/j.pain.2008.06.026>.
- Ramachandran, V. S., and Sandra Blakeslee. *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. Harpercollins Pub Limited, 2005, https://books.google.com/books/about/Phantoms_in_the_Brain.html?hl=&id=RdkfQAAACAAJ.
- Schone, Hunter R., et al. "Making Sense of Phantom Limb Pain." *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 93, no. 8, Aug. 2022, pp. 833–43, <https://doi.org/10.1136/jnnp-2021-328428>.
- Shakiba, Mohamadreza, et al. "Nylon—A Material Introduction and Overview for Biomedical Applications." *Polymers for Advanced Technologies*, vol. 32, no. 9, Sept. 2021, pp. 3368–83, <https://doi.org/10.1002/pat.5372>.
- Talks, Tedx. *Phantom Limbs and Perceived Pain | Jens Foell | TEDxFSU*. Youtube, 22 Sept. 2017, <https://www.youtube.com/watch?v=7-EpZcKgufM>.
- "The Selective Dissolution-Precipitation Recycling: A Review on This Alternative Approach for the Recovery of Nylon-Based Waste Materials." *Waste Management*, vol. 204, Aug. 2025, p. 114937, <https://doi.org/10.1016/j.wasman.2025.114937>.

Understanding Phantom Limb Pain: Causes, Symptoms, and Effective Management. 15 Nov. 2023, <https://minivasivepain.com/phantom-limb-pain/>.

Velasco, Schuyler. "The Future of Prosthetics Is 3D-Printed. This Student Has a Hand in It." Northeastern Global News, 4 Jan. 2024, <https://news.northeastern.edu/2024/01/04/magazine/3d-printed-prosthetic-hands/>.

Wang, Yuxuan, et al. "A Hand Gesture Recognition Strategy Based on Virtual-Dimension Increase of EMG." *Cyborg and Bionic Systems*, Jan. 2024, <https://doi.org/10.34133/cbsystems.0066>.

Website.

https://spiral.imperial.ac.uk/bitstream/10044/1/67183/2/PLP%20Treatment_De%20Nunzio%20AM.pdf.

---. https://www.physio-pedia.com/Mirror_Therapy.

- YouTube. <https://www.youtube.com/watch?v=KdihphPp1Q0>. Accessed 7 Aug. 2025.

---. <https://www.youtube.com/watch?v=chfEJ1USy7s>. Accessed 7 Aug. 2025.

The Impact of Gloomy and Sunny Conditions on AI-Based Rip Current Detection

By Chilok Mak

Abstract

Rip currents are hazardous ocean currents that rapidly flow away from shores and populated beaches, dragging beachgoers out to sea and causing hundreds of drownings annually. This deadly phenomenon highlights the urgent need to both develop reliable detection systems and understand how these classification models behave to improve beach safety. This study explores whether gloomy or sunny conditions contribute to greater computer vision model accuracy in rip current detection. The central research question investigates which weather condition enables the highest model performance, with the hypothesis that sunny conditions lead to higher model accuracy due to greater image visibility. A You Only Look Once (YOLO) convolutional neural network (CNN) was trained on 4,995 labeled coastal images of rip currents and evaluated across test sets representing sunny and gloomy conditions. Contrary to our hypothesis, results showed that the model was more accurate in gloomy conditions than in sunny conditions. These findings challenge assumptions about image clarity by exposing a potential limitation of computer vision systems: reduced reliability in bright, glare-inducing environments. This insight suggests that researchers may need to develop specialized protocols and targeted strategies to address the unique challenges posed by sunny conditions.

Introduction

Rip currents are among the most dangerous natural hazards for beachgoers, responsible for over 100 deaths each year in the United States alone, and tens of thousands of rescues globally [1]. These powerful, narrow currents flow away from the shoreline and can rapidly pull swimmers into deeper water, often leading to panic and drowning [1]. Despite public education and flag systems, many rip currents go undetected or unheeded, especially in conditions where visibility is low or the water appears deceptively calm [1]. Current rip current predictive models typically rely on lifeguard observations, which have human biases and are limited in time and spatial coverage [2]. Accurate, scalable detection methods are essential to reducing fatalities and empowering lifeguards and coastal communities to act preventively.

Artificial intelligence (AI) refers to the simulation of human intelligence by machines programmed to perform tasks such as perception, reasoning, and decision-making [3]. Machine learning, a subfield of AI, enables algorithms to improve their performance on a task through exposure to data without being explicitly programmed for each case [4]. Within this framework, deep learning, specifically convolutional neural networks (CNNs), has proven particularly effective in visual pattern recognition [5]. CNNs are a class of deep neural networks designed to automatically and adaptively learn spatial hierarchies of features through backpropagation using multiple building blocks, such as convolutional layers, pooling layers, and fully connected layers [5]. CNN model architecture is especially suited for analyzing image data, as it excels in tasks such as object detection, classification, and segmentation [5]. CNNs have achieved remarkable

results and are increasingly being applied in environmental monitoring tasks, including oceanographic analysis [6].

Recent studies have focused on utilizing CNNs, computer vision, and other machine learning technologies to implement large-scale rip current detection systems [7-8]. However, before such models can be deployed broadly, it is essential to understand how various environmental conditions, such as lighting and cloud cover, affect detection accuracy. This study addresses that gap, helping inform better dataset construction and model training practices in rip current detection systems.

The objective of our study is to determine whether a computer vision model detects rip currents more accurately under sunny or gloomy conditions. We trained a YOLOv1 CNN-based object detection model on a custom dataset of 4995 annotated rip current images [9-10]. We qualitatively categorized the images as gloomy or sunny based on overcast lighting, and both conditions were equally represented within the dataset (Figures 1 and 2).



Figure 1. Rip current in sunny conditions.

This annotated image was sourced from Roboflow Universe and was part of both the training and testing dataset of the YOLOv1 model.



Figure 2. Rip current in gloomy conditions.

This annotated image was sourced from Roboflow Universe and was part of both the training and testing dataset of the YOLOv1 model.

To ensure our study was robust and grounded in realistic detection scenarios, we drew imagery from a wide range of coastal environments using publicly available beach monitoring sources and open-access video footage. We evaluated the trained model's performance independently on two sets of gloomy and sunny images using metrics such as precision, accuracy, and recall. Each set included an equal number of images with and without rip currents to measure the model's ability to both correctly detect rip currents and reject non-rip currents (Figures 1-4).



Figure 3. Image without rip currents in sunny conditions.

This image was sourced from Roboflow Universe and was part of the testing dataset of the YOLOv1 model.



Figure 4. Image without rip currents in gloomy conditions.

This image was sourced from Roboflow Universe and was part of the testing dataset of the YOLOv1 model.

Our hypothesis predicted that the model would perform better under sunny conditions due to greater visual clarity, higher contrast, and better definition of key image features such as foam channels and breaking waves, which are typically used by human observers to identify rip currents. However, the experimental results challenged this assumption: the model achieved higher accuracy in gloomy conditions than in sunny conditions. This unexpected outcome potentially exposes a limitation in computer vision systems: reduced performance in bright, high-contrast environments. As a result, the study highlights the need for researchers to conduct deeper investigations into the factors that affect accuracy under different lighting and environmental scenarios, paving the way for the development of more resilient and adaptable models.

Materials and Methods

We used an image dataset sourced from Roboflow Universe, a platform that aggregates labeled images from multiple reputable institutions, including the National Oceanic and Atmospheric Administration (NOAA) and the National Weather Service (NWS) [10]. The dataset of 4,995 annotated images of rip currents was split into a 70/20/10 ratio for training, validation, and testing.

We began with a pretrained baseline YOLOv1 CNN and applied transfer learning to adapt it to rip current detection, utilizing the curated training dataset of rip current images [9-10]. The YOLOv1 model's weights were iteratively updated and fine-tuned for 50 epochs of training, with dynamic learning rate adjustments and early stopping to ensure effective generalization to rip current patterns without overfitting.

Once the training was completed, the model was then evaluated on two 1000-image subsets representing gloomy and sunny conditions, with each subset containing 500 images featuring rip currents and 500 without [10-11]. Model evaluation was performed using a custom Python script. The model's predictions on each subset were recorded, with precision, recall, and accuracy calculated and subsequently summarized in bar graph form (Figure 5). Confusion matrices were also generated to further explore classification behavior across true positives, false

positives, true negatives, and false negatives (Figures 6 and 7). These analyses allowed us to assess the effectiveness and reliability of the model under different lighting conditions.

Results

To test our hypothesis that the computer vision model would perform better under sunny conditions, we evaluated its performance on two separate environmental test sets: one composed of sunny beach images and the other consisting of gloomy, overcast scenes. The rationale for this experiment stemmed from the assumption that sunny conditions would offer greater visual clarity, more defined wave features, and higher contrast, all of which are factors thought to support the identification of rip currents by both human observers and automated models.

We trained a computer vision model using a dataset of 4,995 labeled rip current images, optimizing it to detect these hazards by identifying key visual indicators like foam channels, shadowed troughs between sandbars, and breaking wave gaps (Figures 1 and 2). Compiling a testing dataset, we isolated two subsets of images from our full training set. One contained only images captured during sunny weather, and one consisted solely of images captured under gloomy, overcast skies. To assess the model's sensitivity, we added gloomy and sunny images without rip currents to the appropriate subsets [11]. Both environmental sets were balanced in terms of quantity and diversity in beach location, each having 500 images of rip currents and 500 images without rip currents (Figures 1-4).

Evaluating the model's performance on each subset revealed findings contrary to our hypothesis. The model demonstrated slightly higher overall accuracy when analyzing gloomy images compared to sunny ones, with scores of 98.3% and 93.8%, respectively (Figure 5). Beyond overall accuracy, the model was more reliable and consistent on gloomy images than on sunny ones, as shown by precision scores of 99.8% and 93.3%, respectively (Figure 5). Evaluating the model's predictions across all individual images, it failed to detect 16 rip currents and generated 1 false positive in the gloomy set, while it failed to detect 28 rip currents and generated 34 false positives in the sunny set, producing fewer errors in overcast scenes and further indicating superior performance of the model in gloomy conditions (Figures 6-8). Only one set of metrics is presented for each environmental condition since the model produced identical predictions across repeated runs, resulting in no variation between trials.

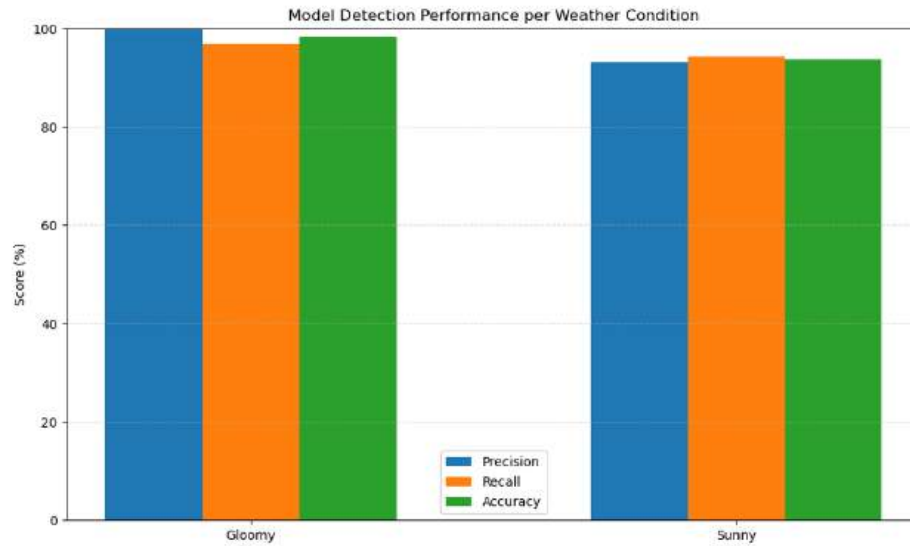


Figure 5. Model accuracy was higher on gloomy images compared to sunny images. Bar graph showing precision, recall, and accuracy of the YOLOv1 model when tested on sunny vs. gloomy image subsets. Metric values were computed and visualized using a Python evaluation script.

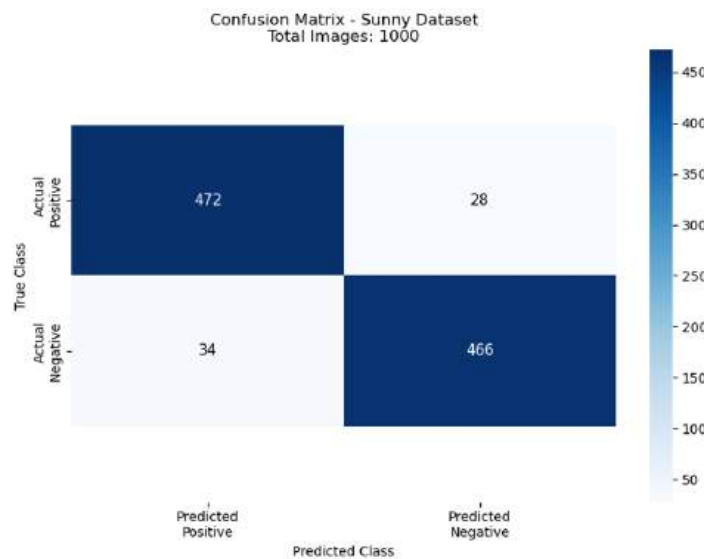


Figure 6. Confusion matrix for the sunny images. Confusion matrix showing true positive, false positive, false negative, and true negative predictions for the YOLOv1 model on sunny test images. Predictions were generated using YOLOv1 weights and evaluated via a custom Python script.

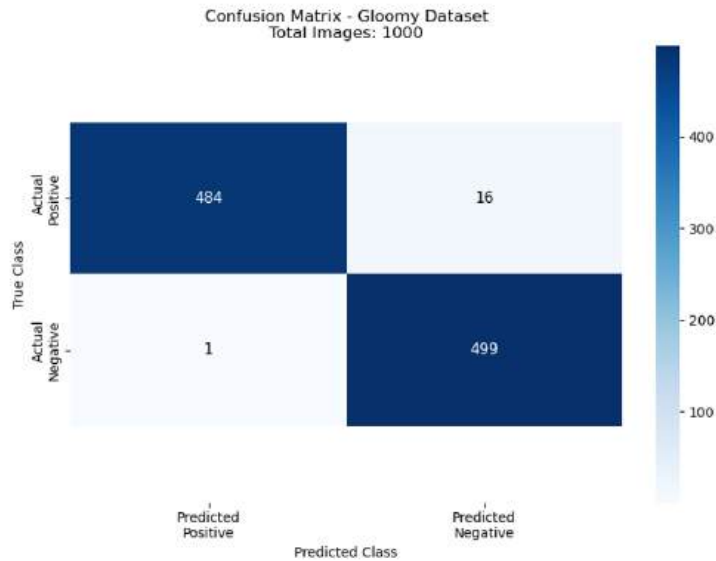


Figure 7. Confusion matrix for the gloomy images.

Confusion matrix showing true positive, false positive, false negative, and true negative predictions for the YOLOv1 model on gloomy test images. Predictions were generated using YOLOv1 weights and evaluated via a custom Python script.

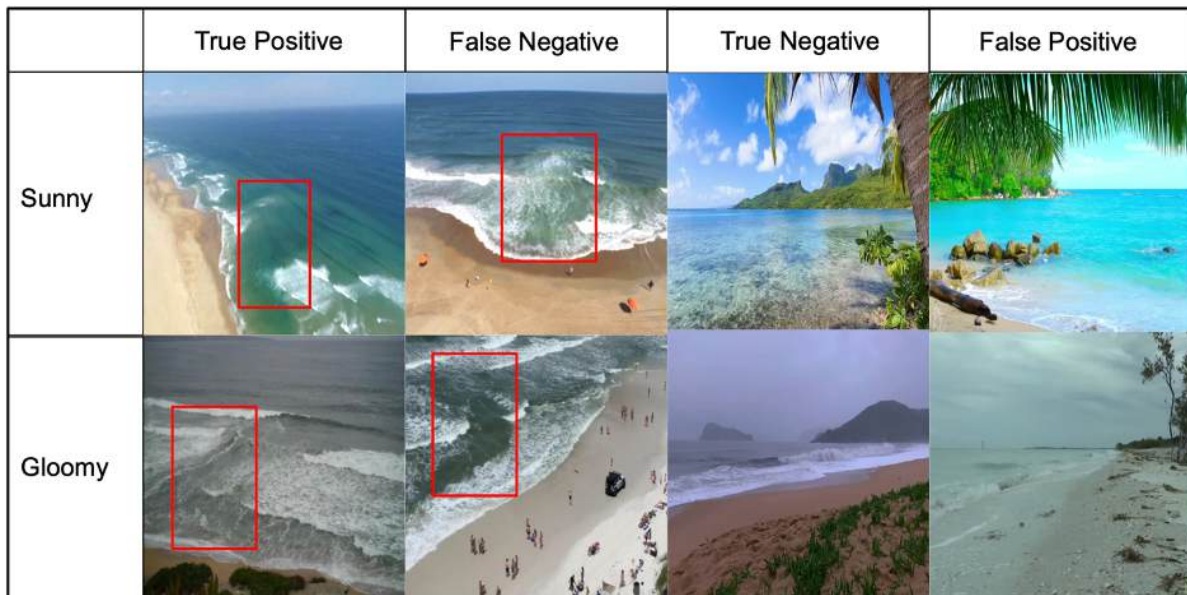


Figure 8. Example images of each classification category.

These images were sourced from Roboflow Universe and were part of the testing dataset of the YOLOv1 model. Predictions were generated using YOLOv1 weights and evaluated via a custom Python script.

Discussion

Automated rip current detection in coastal imagery could provide timely warnings and support lifesaving interventions, as these powerful, fast-moving water channels are a leading cause of surf-related fatalities [12]. Understanding how environmental factors influence model performance is essential for developing more reliable detection systems. This study set out to investigate how lighting conditions affect the accuracy of a computer vision model trained to detect rip currents. Given that high-contrast, brightly lit images are commonly assumed to support better performance in image recognition tasks, we hypothesized that the model would achieve higher accuracy on sunny beach images than on gloomy ones. However, our results suggested otherwise: the model achieved an accuracy of 98.3% on gloomy images, compared to 93.8% on sunny images (Figure 5). These findings challenge our original hypothesis and suggest a more complex relationship between environmental lighting and model performance.

The model's slightly improved performance under gloomy conditions may stem from reduced glare, softer shadows, and more even lighting, which could help the model distinguish key features without interference from bright highlights or harsh reflections commonly found in sunny images. The model's accuracy and precision were slightly higher under overcast conditions, suggesting that diffuse lighting may enhance feature detection, whereas bright sunlight and glare may impair model performance (Figure 5). These findings suggest that the common assumption of brighter conditions yielding better image analysis may not always hold true.

Several factors may have influenced our results. First, the dataset used for training may have contained more consistent or higher-quality examples of rip currents in gloomy images, even if unintentionally. This might have led the model to generalize better under those conditions. Second, the evaluation set may not have evenly represented the diversity within each lighting category. For instance, sunny images may have included a wider range of camera angles and miscellaneous beach objects (e.g., umbrellas and lawn chairs), introducing visual distractions that made predictions more challenging. To reduce this variability and better isolate the effects of lighting, future work could incorporate more images representing each specific condition (i.e., different lighting conditions at the same location with a consistent camera angle), allowing for more precise comparisons across controlled subsets.

These results have several important implications. First, these findings suggest that bright, high-contrast environments, like sunny beaches, may negatively impact the detection performance of computer vision systems. If not properly accounted for, reflective glare may dilute the clarity of visual features, causing the model to misinterpret objects, blur boundaries, or miss subtle patterns that are otherwise critical for accurate detection. This challenges the intuitive idea that clarity and brightness, as understood by humans, correspond to accuracy for computer vision systems. Second, the findings suggest that researchers should recognize changing lighting conditions as an important consideration when evaluating model performance. Rather than assuming standard training methods will generalize across all conditions, it may be necessary to develop specialized systems or preprocessing techniques tailored specifically to handle the

challenges posed by intense lighting. This targeted approach could help improve reliability in real-world settings where such conditions are common.

Ultimately, while this work focused on rip current detection, the findings may hold broader relevance for computer vision in coastal environments. Automated analysis of coastal imagery can support applications including shoreline change detection, ecological surveys, and other beach-related studies [13]. Understanding how lighting conditions shape model performance could therefore inform not only rip current forecasting systems, but also future efforts to develop beach safety and environmental monitoring tools.

Conclusion

This study investigated the impact of environmental lighting on AI-based rip current detection by comparing model performance on sunny and gloomy beach images. While the initial hypothesis predicted that sunny conditions would enhance detection accuracy due to greater visual clarity, the results revealed the opposite: the YOLOv1 convolutional neural network achieved higher accuracy, precision, and overall reliability under gloomy conditions. These findings highlight a potential limitation of computer vision systems in bright, glare-prone environments, where reflections and high contrast may obscure subtle rip current features.

By demonstrating that environmental conditions can meaningfully alter detection outcomes, this research underscores the need for more resilient and adaptive AI models in coastal monitoring. Addressing these challenges may involve developing preprocessing methods to mitigate glare, refining datasets to better balance environmental variability, or training condition-specific models that can adapt to changing beach environments.

Beyond rip current detection, the results hold broader implications for computer vision applications in oceanography and environmental monitoring. As AI becomes increasingly integrated into public safety and scientific observation systems, accounting for environmental variability will be essential to ensure robust, trustworthy performance. Continued research into how lighting, weather, and other natural factors influence detection algorithms will be key to advancing both the accuracy and the real-world applicability of AI-driven beach safety systems.

Works Cited

1. Castelle, B., et al. "Rip current type, circulation and hazard." *Earth Science Reviews*, 163 1-21 (Dec. 2016), www.sciencedirect.com/science/article/pii/S0012825216303117. Accessed 16 August 2025.
2. Dusek, Gregory, and Harvey Seim. "A Probabilistic Rip Current Forecast Model." *ResearchGate* (Jul. 2013), www.researchgate.net/publication/260178309_A_Probabilistic_Rip_Current_Forecast_Model. Accessed 16 August 2025.
3. Russel, Stuart, et al. "Artificial Intelligence: A Modern Approach." *Pearson*. aima.cs.berkeley.edu/newchap00.pdf. Accessed 19 July 2025.
4. Mitchell, Tom. "Machine Learning." *Carnegie Mellon University*. www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf. Accessed 19 July 2025.
5. Rawat, Waseem and Zenghui Wang. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review." *Neural Computation*. doi:10.1162/NECO_a_00990. Accessed 19 July 2025.
6. Wang, Haoyu, and Xiaofeng Li. "DeepBlue: Advanced convolutional neural network applications for ocean remote sensing." *IEEE Xplore Digital Library* (2024), ieeexplore.ieee.org/document/10375549. Accessed 19 July 2025.
7. Dusek, Gregory, et al. "RipScout: Realtime ML-Assisted Rip Current Detection and Automated Data Collection Using UAVs." *IEEE Xplore Digital Library*, vol. 18 (Feb. 2025), ieeexplore.ieee.org/document/10892036. Accessed 19 July 2025.
8. Dusek, Gregory, et al. "RipViz: Finding Rip Currents by Learning Pathline Behavior." *IEEE Xplore Digital Library*, vol. 30, no. 7. (Feb. 2023), ieeexplore.ieee.org/document/10041994. Accessed 19 July 2025.
9. Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Dec. 2016), pp. 779–788. ieeexplore.ieee.org/document/7780460. Accessed 21 July 2025.
10. "Rip_Current_Public_Data Object Detection Dataset." *Roboflow Universe*. https://universe.roboflow.com/sacisro-mwg0q/rip_current_public_data. Accessed 21 July 2025.
11. "Non-rip." *Roboflow Universe*. <https://universe.roboflow.com/advancedrip/non-rip-wcl4c>. Accessed 24 August 2025.
12. Dixon, Richard. "Rip Current Mortality in the Coastal Zone of the United States and its Territories: 2000–2024." *Advances in Hydrology & Meteorology*, vol. 2, no. 5 (Apr. 2025), irispublishers.com/ahm/pdf/AHM.MS.ID.000549.pdf. Accessed 19 July 2025.
13. Tsiakos, Chrysovalantis-Antonios and Christos Chalkias. "Use of Machine Learning and Remote Sensing Techniques for Shoreline Monitoring: A Review of Recent Literature." *MDPI* (Mar. 2023), www.mdpi.com/2076-3417/13/5/3268. Accessed 17 August 2025.

Hierarchical Label Structure-based Curriculum Learning for Lung and Colon Histopathological Image Classification By Junahn Choi

Abstract

Recent advances in deep learning have significantly improved the diagnosis of histopathological images, aiding in the identification of cancer and genetic tissue types. In clinical settings, histopathological image analysis by pathologists incurs high costs, driving interest in AI as a support tool for pathology. However, developing effective pathology AI requires extensive histopathological datasets, which can challenge robust model development. This study addresses these challenges by optimizing training strategies to enhance AI performance on limited datasets. Specifically, we utilized the LC25000 dataset, which contains 25,000 H&E-stained images across five classes, including colon adenocarcinomas, benign colonic tissues, lung adenocarcinomas, lung squamous cell carcinomas, and benign lung tissues. To classify these tissue types, we developed a model employing a curriculum learning approach to improve accuracy under data constraints. A hierarchical label structure was created for colon and lung cancer types, leveraging prior knowledge of disease characteristics to guide model training. This structure enabled the progressive training of the CNN model, starting with easier labels (e.g., Colon vs. Lung tissue) and progressing to more complex labels (e.g., Colon adenocarcinoma vs. Benign colon tissue; Lung disease vs. Benign lung tissue). The proposed training strategy yielded significant improvements in performance, surpassing the baseline model. These results highlight the effectiveness of a progressive training approach for pathology AI with limited data. Our findings suggest that curriculum learning-based training can notably enhance pathology AI performance, offering a cost-effective solution for hospitals seeking reliable AI in histopathological image analysis. Future research may extend this strategy to additional cancer types.

Keywords

Histopathology image classification, Curriculum learning, AI-driven cancer diagnosis, Lung and colon cancer detection

Introduction

Lung and colon cancers are significant public health concerns, ranking among the leading causes of illness and death worldwide. These cancers not only impose an immense burden on healthcare systems but also represent a substantial global challenge in terms of prevention, diagnosis, and treatment. According to recent data from 2024, the United States alone is estimated to see 226,650 new cases of lung cancer and 154,270 new cases of colon and rectal cancer, with approximately 125,070 expected deaths from lung cancer alone (American Cancer Society, *Colorectal Cancer Statistics*; American Cancer Society, *Lung Cancer Statistics*). These statistics underscore the urgent need for effective diagnostic and therapeutic strategies to manage the impact of these cancers.

One critical element of this fight is the early and accurate diagnosis of cancerous cells, which is key to improving patient outcomes and survival rates. Traditionally, cancer diagnosis has relied heavily on histopathology, a process where pathologists closely examine tissue samples under a microscope to identify abnormal cells and determine cancer types. Histopathology remains the gold standard for many cancer diagnoses due to its ability to provide detailed insights into cellular morphology. However, this process is inherently labor-intensive and time-consuming, requiring extensive expertise and manual effort from pathologists. Given the rise in cancer incidence, the high demand for histopathological examination has stretched healthcare resources thin, often leading to delays in diagnosis and treatment.

In recent years, advances in artificial intelligence (AI) and machine learning, particularly deep learning, have revolutionized several fields, including medical imaging and diagnostics. Deep learning models excel in pattern recognition tasks, making them ideally suited for interpreting complex visual data such as histopathological images. These models have demonstrated significant potential in identifying cancerous cells with remarkable accuracy and speed, thus providing a powerful tool for pathologists and healthcare providers (Tolkach et al.; Bulten et al.; Gabralla et al.; Al-Jabbar et al.; Chehade et al.). The automated analysis of histopathological images could drastically reduce diagnostic times, enhance the accuracy of diagnoses, and potentially alleviate some of the workload on pathologists.

Despite this potential, implementing deep learning in histopathology poses challenges. A primary issue is the need for large and diverse datasets to train these models effectively. Real-world histopathological cases vary widely, and creating a model that can generalize well across different tissue samples, staining techniques, and cancer types is challenging without an extensive dataset. Addressing the limitations posed by data scarcity, this study explores the application of curriculum learning, which is a training strategy in machine learning that involves structuring learning in stages of increasing complexity. Curriculum learning has been shown to improve model performance by guiding the learning process, enabling models to develop a more nuanced understanding of complex patterns (Bengio et al.; Lee and Grauman). This approach draws inspiration from human learning principles, where mastering foundational concepts paves the way for tackling progressively challenging tasks.

By structuring our histopathological dataset to incrementally introduce more difficult classification tasks, we hypothesize that curriculum learning can improve the model's generalization capacity, thereby enhancing diagnostic accuracy and robustness. Our study is specifically focused on the classification of lung and colon cancer types within histopathological images. To define the histopathological dataset from easy to hard tasks, we utilized the LC25000 dataset, a well-organized dataset containing histopathological images of five distinct tissue types: colon adenocarcinoma, benign colon tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue (Borkowski et al.). We designed a hierarchical label structure that begins with the simplest task of classifying colon and lung images and progresses to the final classification of all five labels with finer distinctions. This hierarchical setup allowed us to implement a curriculum learning strategy, enabling progressive learning stages that facilitate

model adaptability to increasingly complex tasks. Importantly, this approach required no modifications to the existing model architecture or the addition of external datasets, making it a resource-efficient methodology.

In this study, we demonstrate a 2.47% improvement in diagnostic accuracy, underscoring the effectiveness of our curriculum learning framework for histopathological

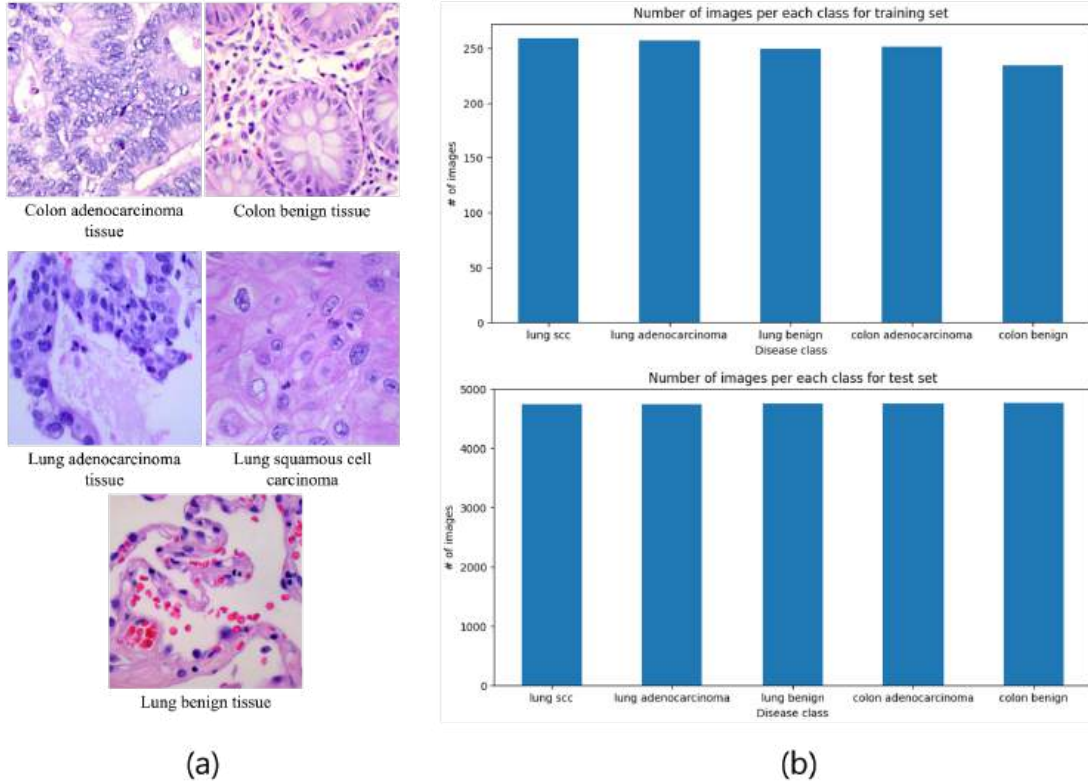


Figure 1. (a) Visualization of colon and lung tissues from LC25000 dataset and (b) Number of images per each disease class for training and test set.

image classification. This work highlights the potential of curriculum learning to enhance cancer diagnostics, enabling more efficient and reliable detection while addressing current limitations in data availability and model robustness.

Methodology

Description of Dataset: For training the Convolutional Neural Network (CNN) model for histopathological diagnosis, we used the LC25000 dataset from Kaggle databases, a widely recognized resource containing a substantial collection of histopathological images used for diagnostic research (Borkowski et al.). The LC25000 dataset provides a total of 25,000 hematoxylin and eosin (H&E)-stained images of colon and lung tissues, with five distinct classes. These classes are categorized as colon adenocarcinoma, benign colon tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. The LC25000 dataset was generated by augmenting an initial set of 1,250 unique images. Through techniques such as

rotations, translations, and scaling, the dataset expanded to 25,000 samples. This augmentation process aimed to create a larger, more diverse dataset, improving the model's generalization capability and robustness across different tissue presentations.

The H&E staining technique, a standard in histopathology, uses hematoxylin to stain nuclei blue and eosin to stain cytoplasmic and extracellular components pink. This color differentiation enhances cellular details and structural distinctions, allowing for more effective diagnosis and classification of tissue samples. Within the LC25000 dataset, these characteristics help distinguish each disease type visually. For instance, adenocarcinomas,

Disease type	Classification label		
	<i>Stage 1</i>	<i>Stage 2</i>	<i>Stage 3</i>
Colon benign	C1	C1	C1
Colon adenocarcinoma	C1	C2	C2
Lung benign	C2	C3	C3
Lung adenocarcinoma	C2	C4	C4
Lung squamous cell carcinoma	C2	C4	C5

Table 1. Hierarchical label structure of diseases for histopathological image. C refers to class for each stage.

both in colon and lung, display abnormal glandular structures and prominent nuclei, indicating malignant activity, while benign tissue samples retain organized cellular architecture without atypical nuclei. Lung squamous cell carcinoma typically exhibits polygonal cells with prominent keratinization and intercellular bridges, which are observable in the squamous differentiation.

Due to the heavy augmentation applied to the LC25000 dataset, we strategically divided it into 1,250 images for the training set and 23,750 for the test set. This split minimizes the overlap between training and testing data, thereby reducing potential bias and ensuring that model evaluation is based on images that were not seen during training. By carefully curating the training and test sets in this manner, our study aims to establish a more rigorous and unbiased benchmark for evaluating model accuracy and generalizability in histopathological image classification. Figure 1 (a) provides visual examples for the five tissue types included in the LC25000 dataset. Figure 1 (b) visualizes the dataset distribution across each disease class for training and test sets.

Designing Hierarchical Label Structure: In order to implement an effective curriculum learning strategy, this study introduces a hierarchical label structure that allows the Convolutional Neural Network (CNN) model to progressively learn the task of histopathological image classification. This structure was specifically designed to handle the challenge of defining task difficulty in a systematic way, progressing from simpler tasks to more complex ones.

As shown in Table 1, the hierarchical structure is divided into three stages based on the complexity of the disease classification tasks:

Stage 1 – Organ-Level Classification. In the first stage, the model is tasked with a binary classification, where it only distinguishes between images of colon and lung tissues. This introductory level reduces the complexity of the task, providing the model with an initial, broad differentiation between the two major organ types. This setup allows the model to establish foundational discrimination skills between colon and lung tissues, focusing on broader morphological features rather than the specific characteristics of malignant or benign samples.

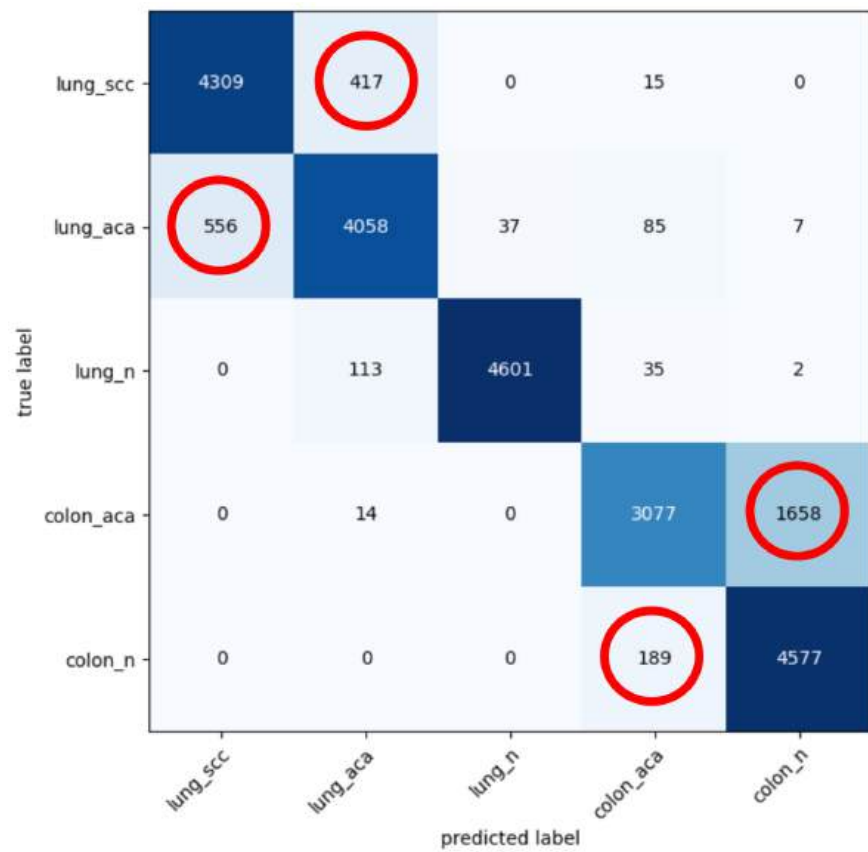


Figure 2. Confusion matrix of baseline model for fine-grained disease classes. The red circle highlights the cases of difficult classification.

Stage 2 – Benign vs. Malignant Identification within Organs. After establishing the organ-level classification, the model proceeds to Stage 2, which introduces additional complexity by differentiating benign from malignant tissues within each organ type. This stage involves four classes: benign colon, colon adenocarcinoma, benign lung, and malignant lung tissue. At this stage, the model begins to identify specific disease states within each organ type, moving from broad anatomical identification to understanding basic pathological distinctions. The stage builds upon the knowledge acquired in Stage 1, adding another layer of discrimination based on benign and malignant features.

Stage 3 – Fine-Grained Classification Across All Classes. The final stage in the hierarchy, Stage 3, requires the model to distinguish between five specific classes of tissues: benign colon, colon adenocarcinoma, benign lung, lung adenocarcinoma, and lung squamous cell carcinoma. At this point, the model is expected to make more nuanced distinctions within malignant categories, such as differentiating lung adenocarcinoma from lung squamous cell carcinoma. This classification level represents the highest complexity, as the model must rely on intricate morphological features to classify cancer subtypes accurately.

To explore the rationale behind designing a hierarchical label structure, which progresses through organ-level classification, malignant identification, and fine-grained disease classes, we conducted an experiment using the baseline model trained directly on the original fine-grained disease classes. The training process was carried out for one hundred epochs using cross-entropy loss. As depicted in Figure 2, the results highlighted several challenging cases in classifying colon and lung histopathological images. Specifically, the model struggled to differentiate between normal and adenocarcinoma tissues in colon images, as well as between adenocarcinoma and squamous cell carcinoma in lung images. These difficulties stemmed from overlapping morphological features and subtle variations that are challenging to capture in a direct, fine-grained classification task.

To address these challenges, we designed the hierarchical label structure based on the observed classification difficulties. The staged approach allows the model to start with broader tasks, such as distinguishing between colon and lung tissues, before progressing to more complex tasks like identifying benign versus malignant tissues and fine-grained disease subtypes. This design enables the model to incrementally learn and refine its features, effectively tackling the challenges of histopathological image classification.

Model Architecture: The proposed convolutional neural network (CNN) efficiently extracts hierarchical features through a series of convolutional, activation, and pooling layers. It begins with two convolutional layers (32 and 64 filters) followed by max pooling, progressively reducing spatial dimensions. Deeper layers include convolutional blocks with 128 and 256 filters, culminating in another max-pooling layer to refine features further. The model has 388,416 trainable parameters with an efficient memory of 4.99 MB.

Overall Training Strategy: To effectively implement the hierarchical label structure, the model training process is designed as a sequential curriculum learning strategy, wherein the CNN progressively learns the task by transitioning through three stages of increasing classification complexity. This method ensures that the model builds a strong foundation before addressing more nuanced and challenging distinctions in histopathological image classification. The training strategy for three stages are as follows.

Stage 1 – Organ-level training. At the first stage, the model undergoes a supervised training process focused solely on binary classification at the organ level. The objective is to differentiate between colon and lung tissues based on their broad morphological characteristics.

The training data at this stage is balanced to avoid bias toward either organ type. By focusing on simpler tasks, this stage enables the model to develop core discriminatory features, such as organ-level morphological patterns, without being overwhelmed by pathological complexities.

Stage 2 – Benign vs. Malignant differentiation within organs. After achieving satisfactory performance in Stage 1, the model progresses to Stage 2, where the complexity is increased by introducing four classes: benign colon, colon adenocarcinoma, benign lung, and malignant lung tissues. The weights learned in Stage 1 are retained and fine-tuned to accommodate the new task. By leveraging the foundational organ-level discrimination skills acquired earlier, the model is trained to recognize pathological differences within each organ type. This stage emphasizes learning the morphological features that distinguish benign from malignant samples while still maintaining organ-specific context.

Stage 3 – Fine-grained cancer subtype classification. In the final stage, the model is exposed to all five classes: benign colon, colon adenocarcinoma, benign lung, lung adenocarcinoma, and lung squamous cell carcinoma. This stage requires the model to learn subtle morphological differences between various cancer subtypes, particularly between lung

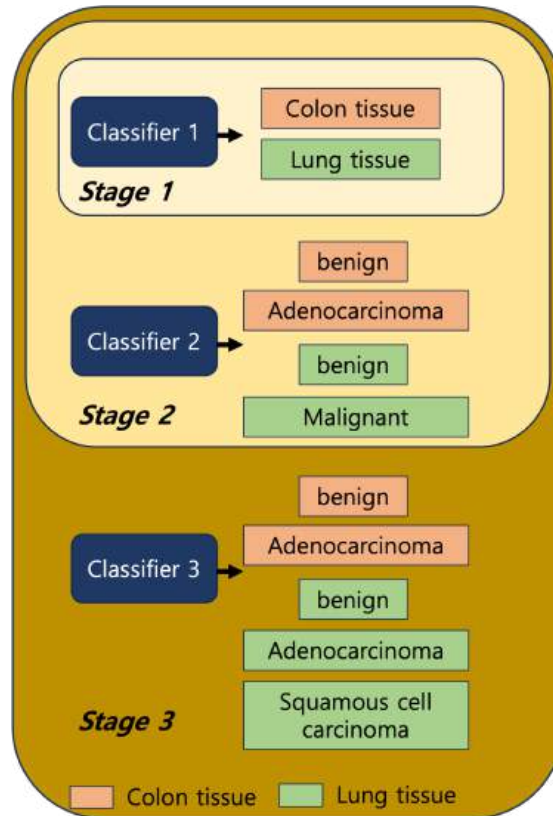


Figure 3. A training pipeline with three stages for hierarchically labeled structure-based curriculum learning.

adenocarcinoma and lung squamous cell carcinoma, which often exhibit overlapping features. Transfer learning from the previous stages ensures that the model builds upon its existing knowledge, reducing the computational burden and facilitating more efficient training. To ensure

continuity and effective knowledge transfer across stages, the training strategy incorporates prior classification loss alongside the current stage loss. This approach allows the model to retain and refine previously learned features while adapting to the increased complexity of subsequent stages. By integrating past and present learning objectives, the model achieves a balanced progression, enhancing its ability to generalize across hierarchical tasks in histopathological image classification.

Training Details and Evaluation metrics: The CNN model undergoes training for a total of 150 epochs, divided into three distinct stages to ensure stable and effective convergence. In Stage 1, the model is trained for fifteen epochs to initialize the learning process, focusing on establishing a foundational understanding of the data. Stage 2, also lasting fifteen epochs, aims to refine the features learned during the initial phase by progressively adapting to more complex patterns in the dataset. Finally, Stage 3 spans one hundred and twenty epochs, allowing the model to further fine-tune its parameters for better performance.

Method	Evaluation metrics			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>
Baseline	89.59 %	0.907	0.896	0.894
Ours	92.06%	0.921	0.921	0.920

Table 2. Classification comparison between baseline and our method using hierarchical label structure-based curriculum learning.

For optimization, the cross-entropy loss function is utilized to measure the performance of the model during training by comparing its predictions with the ground truth labels. The Adam optimizer, configured with a learning rate of $1e-3$, is employed to update the model's parameters efficiently and robustly, ensuring faster convergence while maintaining stability (Kingma). To improve the model's ability to generalize across diverse data distributions, several data augmentation techniques are applied. These include resizing all input images to a uniform size of 128×128 pixels to standardize the dataset and applying random horizontal flips to artificially increase the dataset's variability. These augmentation strategies help the model better adapt to variations in the input data, reducing the risk of overfitting and enhancing its robustness.

For evaluation, the metrics used include accuracy, precision, recall, and F1 score, each serving distinct purposes in assessing the model's performance.

Accuracy: Measures the overall correctness of the model's predictions. Accuracy is formulated as $(TP + TN) / (TP + TN + FP + FN)$, where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

Precision: Focuses on the quality of positive predictions by determining the ratio of correctly predicted positive samples to all predicted positives. It is formulated as $TP / (TP + FP)$.

Recall: Emphasizes the ability of the model to correctly identify all actual positive instances. It is formulated as $TP / (TP + FN)$.

F1 Score: Provides a balance between Precision and Recall by calculating their harmonic mean. F1 Score is formulated as $2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall})$.

Experimental results

We evaluated the performance of baseline and our proposed method over four evaluation metrics. By leveraging the hierarchical label organization on the LC25000 dataset, which incorporates prior knowledge of colon and lung cancers, we achieved a substantial improvement in the classification of histopathological images. As shown in Table 2, our method improved overall performance by 2.47% in accuracy, 0.014 in precision, 0.025 in recall, and 0.026 in F1 score, respectively.

In the confusion matrices shown in Fig. 6, our proposed hierarchical label structure-based curriculum learning performs better and more balanced in classifying the challenging cases, particularly between lung adenocarcinoma and lung squamous cell carcinoma, as well as between colon benign and adenocarcinoma. Specifically, the baseline model exhibits a bias toward classifying samples as colon benign. In contrast, our proposed

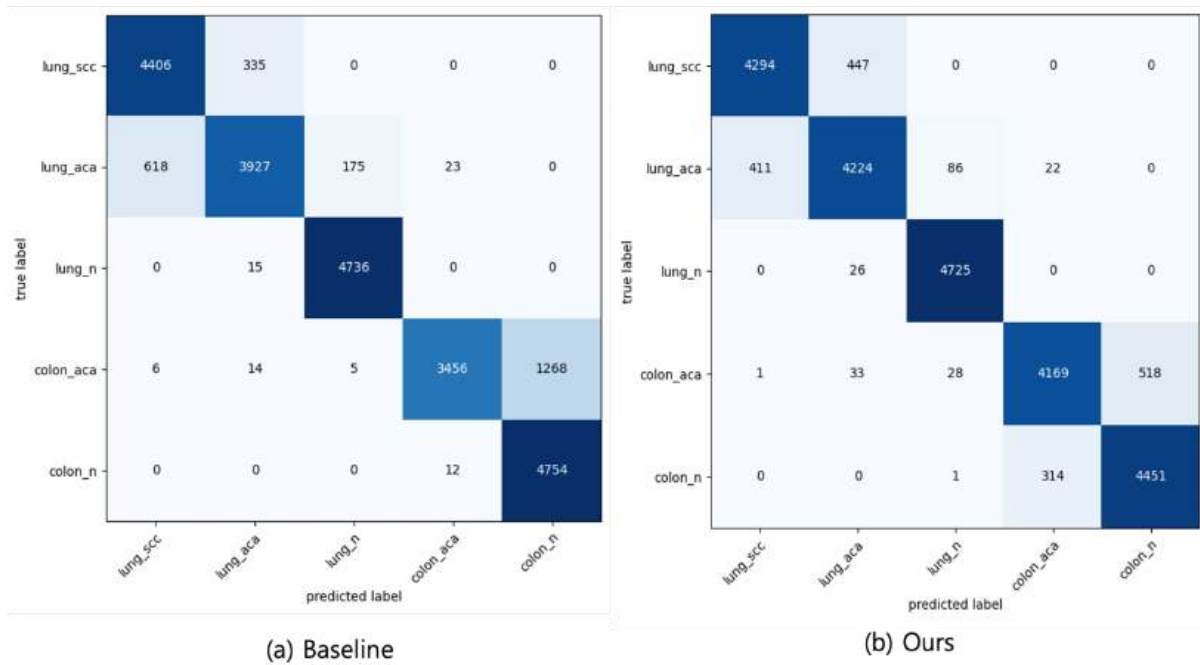


Figure 4. Confusion matrices of the baseline model and our proposed method on the LC25000 test set. Here, n, aca, and scc refer to benign, adenocarcinoma, and squamous cell carcinoma, respectively.

method reliably distinguishes colon cancer, demonstrating its stability and effectiveness in handling these classifications. Through these results, we validated the effectiveness of our proposed method in improving classification performance for challenging histopathological image tasks.

Discussion

The training strategy proposed in this study provides a simple yet effective method for improving multi-class classification by utilizing prior knowledge of diseases. By introducing a hierarchical label structure, the method simplifies the learning process, enabling the model to progressively tackle increasingly complex classification tasks. This approach successfully enhances classification performance without requiring additional datasets or increasing the complexity of the model during inference.

Multi-class classification tasks often involve inherent relationships among classes, as observed in various domains like medical imaging, biology, and taxonomy. Organizing these relationships into a hierarchical structure allows for a systematic application of curriculum learning, where tasks are ordered by difficulty. By defining labels as easy or hard based on disease characteristics, the model gains the ability to adapt gradually to more challenging distinctions, improving its overall performance and generalization.

However, this strategy comes with notable challenges. A significant limitation is the reliance on domain knowledge to define task difficulty and structure the hierarchical relationships. Such knowledge is not always universally available or consistently well-defined, which can limit the applicability of the approach across different datasets or domains. Additionally, determining accurate difficulty levels within the curriculum learning framework remains a complex and subjective task, which can directly impact the model's effectiveness.

Despite these challenges, the findings of this study highlight the practical potential of curriculum learning when applied to medical image classification. The simplicity and accessibility of the proposed method make it suitable for a wide range of applications, while its reliance on prior knowledge underscores the importance of domain expertise in advancing machine learning solutions. Future research can address the identified limitations by expanding datasets, adopting more complex models, and exploring automated techniques for defining hierarchical relationships, thereby broadening the scope and applicability of this approach.

Conclusion

Using prior knowledge of diseases, we proposed a simple and efficient training strategy that is three stage hierarchical label structure-based curriculum learning for lung and colon histopathological image classification. We can easily organize the hierarchical label structure and successfully adopt the curriculum learning. By systematically lowering learning difficulties, we achieved notable performance improvements, including a 2.47% increase in diagnostic accuracy, without requiring additional datasets. Finally, our proposed training strategy would be expected to be effective for certain scenarios which are multi-class classification and have hierarchical structure relation among labels.

Works Cited

- Al-Jabbar, M., Mohammed Alshahrani, Essam M. Senan, and Ibrahim A. Ahmed. "Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features." *Bioengineering*, vol. 10, no. 383, 2023.
- American Cancer Society. Key Statistics for Colorectal Cancer. Cancer.net, 2024, <https://www.cancer.net/cancer-types/colorectal-cancer/statistics>.
- American Cancer Society. Key Statistics for Colorectal Cancer. Cancer.net, 2024, <https://www.cancer.net/cancer-types/colorectal-cancer/statistics>.
- Bengio, Yoshua, Jérôme Louradour, Ronan Collobert, and Jason Weston. "Curriculum Learning." *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 41–48.
- Borkowski, Andrew A., Minh M. Bui, Laura B. Thomas, Charles P. Wilson, Lisa A. DeLand, and Stephen M. Mastorides. Lung and Colon Cancer Histopathological Image Dataset (LC25000). arXiv:1912.12142, 2019, <https://arxiv.org/abs/1912.12142>.
- Bulten, Wouter, Karlis Kartasalo, P.-H. C. Chen, Petter Ström, Henk Pinckaers, Kushal Nagpal, Yong Cai, Daniel F. Steiner, Henk Van Boven, Rob Vink, et al. "Artificial Intelligence for Diagnosis and Gleason Grading of Prostate Cancer: The PANDA Challenge." *Nature Medicine*, vol. 28, 2022, pp. 154–163.
- Chehade, Ahmad H., Nader Abdallah, Jean-Marie Marion, Mohammad Oueidat, and Philippe Chauvet. "Lung and Colon Cancer Classification Using Medical Imaging: A Feature Engineering Approach." *Physical and Engineering Sciences in Medicine*, vol. 45, 2022, pp. 729–746.
- Gabralla, Lina A., Ahmed M. Hussien, Abdulrahman AlMohimeed, Hadeel Saleh, Doaa M. Alsekait, Samir El-Sappagh, Abdelrahman A. Ali, and Mohammed R. Hassan. "Automated Diagnosis for Colon Cancer Diseases Using Stacking Transformer Models and Explainable Artificial Intelligence." *Diagnostics*, vol. 13, no. 2939, 2023.
- Kingma, Diederik P. "Adam: A Method for Stochastic Optimization." arXiv:1412.6980, 2014, <https://arxiv.org/abs/1412.6980>.
- Lee, Yong Jae, and Kristen Grauman. "Learning the Easy Things First: Self-Paced Visual Category Discovery." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1721–1728.
- Tolkach, Y., T. Dohmgörge, M. Toma, and G. Kristiansen. "High-Accuracy Prostate Cancer Pathology Using Deep Learning." *Nature Machine Intelligence*, vol. 2, 2020, pp. 411–418.

Domination and Disavowal: Israel, Palestine, and the Question of the Settler-Colonial Framework By Ariya Harrington

“As observed, the logic of elimination continues into the present.”—Patrick Wolfe

Abstract

This literature review examines the emerging prominence of the settler-colonial paradigm within the field of Israeli-Palestinian history. It analyzes twenty-first-century scholarly publications, with an eye toward the events of October 7th, 2023, and the subsequent growth of mainstream attention to the ongoing War in Gaza. I acknowledge the contentious nature of the discussion surrounding settler colonialism and Palestine, then move into a brief overview of the history of Palestine and the formation of the state of Israel. I then progress to analyzing the works of scholars arguing the necessity of viewing Israel-Palestine through a settler-colonial lens, and attempt to deconstruct the logic of their arguments and their perspectives. Finally, I touch on alternate perspectives, examining the counternarratives that have emerged in conjunction with this trend which argue against the use of the settler-colonial lens within Palestine studies. Finally, I assess the more recent growth of popular interest in the term “settler colonialism” in the wake of the October 7th attacks and Israel’s military invasion of the Gaza Strip.

Introduction

The issue of Israel and Palestine remains a contentious and hotly debated topic in politics, media, and scholarship alike. Consequently, the question of whether Israel is best viewed through the lens of “settler colonialism” has also emerged as a complex and controversial academic battleground. Discussions of the conflict on digital platforms, though frequent and open to vast, unfiltered participation, often exclude decades of history and reduce valuable context to isolated events.

To better grasp the question of the settler-colonial label, it becomes necessary not only to acknowledge the origins of the Israel-Palestine conflict as they emerged and evolved over the course of the twentieth century, but also to define “settler colonialism” as a compound term distinct from the more traditional label of “colonialism.” “Colonialism” is used to describe systems of colonization dependent on the extraction of resources from a colony, often resulting in the exploitation of indigenous populations by the colonizing power (Veracini 2013, 317). The British Raj in India and the Belgian Congo are commonly cited as examples of explicit colonial domination over native peoples. “Settler colonialism” is distinguished as a process of colonization with a particularly central focus on land, its obtainment, and the consequent intent to remove, eliminate, displace, or destabilize the native population (Veracini 2013, 319). Many scholars describe Australia and the United States as examples of settler colonies, massively displacing and erasing existing indigenous peoples.

In recent years, scholarship is shifting toward situating and understanding the issue of Palestine and Israel within the broader settler-colonial framework; specifically, some scholars argue in favor of this shift in order to facilitate a more globalized comparison of the Palestine situation. Notably, a number of scholars writing on Palestine have taken inspiration from or built directly on contributions made to the field of settler-colonial studies by the late anthropologist Patrick Wolfe. With an eye toward the events of October 7th, 2023, this literature review seeks to examine the genealogy of the contemporary settler-colonial paradigm within scholarly contributions to Palestine studies in recent years, analyzing the arguments made by scholars while simultaneously deconstructing their distinct justifications and perspectives.

Historical Overview: Contextualizing the Issue of Israel-Palestine

The origins of the conflict in Palestine can be traced back as far as the late 19th century to the earliest waves of Jewish immigration. The arrival of the earliest Zionist colonists in 1882 marked the First Aliya (a Hebrew term indicating immigration to Israel), which ended in 1903 and saw the immigration of approximately 20–30,000 European Jews. The Second Aliya would span from 1904–1914, and was characterized by the immigration of about 35–40,000 Jews to Palestine (Shafir 1999, 76). The resonance of the historic Jewish connection to the land of Israel, understood by many Jews as the “Promised Land” for over 3,000 years, strongly encouraged migration to the region. During the early 20th century, Zionist leaders made the colonial intentions of immigration explicit (for example, the Jewish Colonization Association); it was during the period of decolonization and the depopularization of colonial attitudes that these practices were rebranded and/or erased (Khalidi 2020, 17).

In 1917, following the collapse of the Ottoman Empire during World War I, and the end of Ottoman control, Britain pledged its support for the formation of Jewish statehood and sovereignty in Palestine through the Balfour Declaration; it was in this agreement that the existing Palestinian Muslim and Christian population was mentioned as simply “non-Jewish,” and left unaccounted for in the discussion of national rights (Khalidi 2020, 15).

Just five years later, the League of Nations built on the Declaration via the Mandate for Palestine, which formalized British governance over Palestine. Though the Mandate facilitated Jewish immigration and settlement, it was only after the rise of the Nazis and the intensification of Jewish persecution in Germany that Jews were driven from Germany to Palestine in mass numbers. Subsequently, and for the first time, the Jewish economy (largely segregated from that of the Arabs) overtook the Arab sector (Khalidi 2020, 43).

It was in 1937 that the British Royal Commission proposed an official partition of Palestine, which would create a Jewish state in approximately 17% of the existing Palestinian territory. Enraged, many Arabs responded via armed revolt, intensifying to the point where rebels successfully claimed control of much of the Palestinian countryside, as well as several urban areas. However, the British brutally suppressed and put an end to the revolt by 1939 using executions, home demolitions, and other violent military tactics in their counter-insurgency program (Khalidi 2020, 46-47).

After the *Yishuv* (the Jewish population living in Palestine before the founding of Israel) drove the British from Palestine in the late 1940s, the Palestinian issue was placed under the jurisdiction of the United Nations' Special Commission on Palestine. The dominant powers, in a shift better anticipated by Zionists than Arabs, had become the U.S. and the Soviets. Resolution 181 in 1947 officially divided Palestine into a larger Jewish state (over 56% of the land) and a smaller Arab state. Arab leaders outright rejected the partition plan, as it gave not only a majority of the territory, but also the best agricultural lands to the recently arrived European Jews (Khalidi 2020, 72).

The first phase of what Palestinians deemed the “Nakba,” or catastrophe, began in late 1947 and was characterized by the conquest and depopulation of the territory's largest urban centers. After the State of Israel was officially founded on May 15, 1948, the Nakba continued into its second phase. The Israeli army defeated the Arab armies that joined during the war, and massive numbers of Palestinians were expelled from Israeli territory to neighboring states such as Jordan, Syria, Lebanon, as well as the West Bank and Gaza. As the new Jewish state was established, a total of around 750,000 Palestinians were displaced as refugees to surrounding Arab countries (Baig 2019, 52).

Violent interactions between Israelis and Palestinians would only continue in the following decades. The Suez War in 1956 saw Britain and France join Israel's invasion of Egypt to suppress Egyptian nationalism, support for Algerian rebels, and nationalist challenges to British imperialism in other Arab regions. During this period, numerous massacres of Gazans were carried out, and Israel occupied the Gaza Strip until the United States and the USSR forced an end to the tripartite occupation of both Egyptian and Gazan territory (Khalidi 2020, 92).

Less than a decade later, Israel occupied Palestinian territories—the West Bank, East Jerusalem, and Gaza—once more during the June 1967 Arab-Israeli War. The ambiguous wording of UN Security Council Resolution 242 in 1967 called for the “withdrawal of Israeli armed forces from territories occupied in the recent conflict,” effectively legitimating the 1949 armistice lines (also known as the Green Line) as Israel's boundaries (Khalidi 2020, 105). This resolution marked a significant resurgence of Palestinian national consciousness, identity, and resistance in response (Khalidi 2020).

In 1982, the Israeli invasion of Lebanon marked the beginning of the Lebanon War. The invasion targeted the Palestine Liberation Organization's base of operations, which had by then built up its military and political strength, to reduce the capacity of Palestinian resistance and the nationalist movement. The First Intifada broke out just five years later across the occupied territories, following two decades of Israeli occupation of the West Bank and the Gaza Strip (Khalidi 2020, 161). Although the First Intifada initially bolstered Palestinian national identity and improved Palestinian image on the international stage, the Second Intifada in 2000 was marked by extensive suicide bombing assaults carried out inside Israel by groups such as Hamas, Islamic Jihad, Fatah, and other factions in the PLO. The Second Intifada was ultimately a major setback for the Palestinian national movement, and culminated in the intensification of Israel's occupation of the West Bank and the Gaza Strip (Khalidi 2020, 198).

Following the 1982 war, historiography within Israeli scholarship underwent a significant shift. In their publications, Israel's self-designated "New Historians" reported on new material found primarily in declassified archival documents, re-contextualizing Israel's history and challenging the foundations of Zionist scholarship. Israeli sociologist Gershon Shafir's publication, "Zionism and Colonialism," was groundbreaking in the connections it drew between Zionist ideology and colonial practices (Shafir, 1999). Historian Gabriel Piterberg notes that although the shift in scholarly perspective within Israel was a fundamental change, what the "New Historians" offered was "neither a new historiography nor a radical critique of Zionism as ideology and praxis" (Piterberg 2015, 19). Piterberg also acknowledges that this shift was facilitated more by Israeli historical sociologists than by historians, due to general disagreement over the exceptionalism of the Zionist project (Piterberg 2015, 19). Gershon Shafir's intervention would have a crucial impact on the trajectory of contemporary settler-colonial theory as it developed under Patrick Wolfe and Lorenzo Veracini.

The Growing Emergence of the Settler-Colonial Framework

Australian anthropologist Patrick Wolfe is widely credited with the popularization of settler-colonial studies in its current form. Wolfe's 2006 article "Settler Colonialism and the Elimination of the Native," published in the *Journal of Genocide Research*, examines settler colonialism not as a singular event, but a complex land-centric structure of invasion following a "logic of elimination." This "logic" dictates that a settler-colonial power is dependent on the destruction or "elimination" of native societies and their ties to the land in an invasive process often coinciding with genocide (Wolfe 2006, 388). Using this reasoning, he examines a number of settler-colonial societies including those in Australia, North America, and Palestine, and subsequently categorizes the removal, assimilation, and genocide of indigenous populations as strategies used to facilitate elimination. In his explanation of settler colonialism as a force which "destroys to replace," Wolfe cites Israel as a settler colony. Wolfe consistently returns to Palestinians as an example of a native society subject to elimination by Israeli settler society, stating, for example, that "as the logic of elimination has taken on a variety of forms in other settler-colonial situations, so, in Israel, the continuing tendency to Palestinian expulsion has not been limited to the unelaborated exercise of force" (Wolfe 2006, 401).

The 2010s saw several critical works published on the subject of settler colonialism in Palestine. The article "Past is Present: Settler Colonialism in Palestine" emerged from a 2011 conference on settler colonialism in Palestine and was published in 2012 as a special issue in the *Settler Colonial Studies Journal*. Salamanca et al. (2012) builds on Patrick Wolfe's "logic of elimination," arguing that the threat of "[physical and symbolic] elimination" of native populations necessitates realigning the goals of scholarship and liberation movements with decolonization in order to push the field of settler-colonial studies into generating tangible action rather than conjecture (Salamanca et al. 2012, 4). Furthermore, "Past is Present" asserts that establishing a more widespread use of "settler colonialism as a relevant interpretative framework for the study of Zionism" allows for comparison between Israel and other settler-colonial structures, sidestepping the interpretation of Israeli exceptionalism as well as empowering and

unifying more globalized struggles against colonialism, such as in South Africa or French Algeria (Salamanca et al. 2012, 4). The article argues that the modern history of Palestine must be viewed through the lens of settler colonialism to best understand the ideologies and actions of the Israeli state, and to compare the issue of Israel-Palestine within a larger and more globalized context of colonial history.

Published just a year later, Lorenzo Veracini's 2013 article "'Settler Colonialism': Career of a Concept" examines how settler colonialism has evolved as a distinct category of historical analysis. "Career of a Concept" explores how "settler colonialism" only began its emergence as a compound term in the context of later 20th-century anti-colonial struggles. As settler colonialism became more widely understood by scholars as a structure relating both settler domination and existing native society (a process which had not previously been embraced), ethnohistory, anthropology, indigenous studies, and history developed into more crucially interrelated fields of study (Veracini 2013, 324). Veracini's (2013) work built significantly on the existing contributions of Patrick Wolfe; though Veracini (2013) has undoubtedly contributed much to the field of settler-colonial studies, he has credited Wolfe for his influence. Unlike "Past is Present," "Career of a Concept" was not written with a focus on Palestine. However, it does ascertain that "the very acknowledgement of a settler colonial dimension" applied to the modern Palestine situation "is bound to be most contentious," demonstrating an awareness of the significance of settler-colonial studies to the issue of Palestine (Veracini 2013, 314).

Historian Gabriel Piterberg's 2015 article "Israeli Sociology's Young Hegelian: Gershon Shafir and the Settler-Colonial Framework" was published through the *Journal of Palestine Studies*. It examines the evolution and impact of Gershon Shafir's work within the field of Israeli sociology, reviewing the literature of the New Israeli Historians, Gershon Shafir's own works, and more recent works bearing Shafir's influence. Piterberg (2015) makes the argument that Shafir was the first Israeli sociologist to break away from established norms by bridging the gap between Zionist colonization in Palestine and broader comparative studies of settler colonialism. Similar to Salamanca et al. (2012), Piterberg (2015) emphasizes the importance of comparing the issue of Israel and Palestine to other settler-colonial situations across history and geography, which he believes can only be achieved through bringing settler-colonial studies to greater prominence within the field of Palestine studies. He posits that Wolfe and Veracini's publications were both heavily influenced by Shafir's work *Land, Labor, and Origins*, which he claims influenced him significantly as well.

Rashid Khalidi's 2020 book *The Hundred Years' War on Palestine* is an account of the Israel-Palestine conflict, tracing the history of the region from the late 1800s to the present day. It must be noted that Khalidi's (2020) account is distinct from previously examined publications in this review in that it takes a more indigenous-centered approach to the chronology of escalation and resistance, grounding Khalidi's perspective firmly in his experiences as a Palestinian man personally present for several of the events he recounts. Khalidi (2020) argues plainly in favor of understanding Palestinian history "as a colonial war waged against the indigenous population, by a variety of parties, to force them to relinquish their homeland to another people against their

will” (Khalidi 2020, 13). Similarly to Wolfe, Khalidi (2020) cites Theodore Herzl’s justifications for Zionism as a “touchstone for colonialists” and a crucial element of the ongoing argument in favor of the Zionist ideology (Khalidi 2020, 10). Indeed, Khalidi (2020) seems to at least partially corroborate Wolfe’s “logic of elimination,” claiming that, “settler-colonial confrontations with indigenous peoples have only ended in one of three ways: with the elimination or full subjugation of the native population...with the defeat and expulsion of the colonizer...or with the abandonment of colonial supremacy” (Khalidi 2020, 231).

Counternarratives and Alternate Perspectives

The field of settler colonialism continues to move closer toward mainstream discussion of Palestinian and Israeli history. Nevertheless, counternarratives have emerged in conjunction with this trend, with some arguing against the use of the settler-colonial label to describe Israel.

Alan Dowty, Professor Emeritus of Political Science at the University of Notre Dame and a visiting scholar at the Stroum Center for Jewish Studies, published his 2022 article “Is Israel a Settler Colonial State?” through the Stroum Center for Jewish Studies. Formerly the president of the Association for Israel Studies, Dowty argues that it is not appropriate to define the state of Israel as a product of “settler colonialism.” Though he concedes that one may fairly characterize early Zionist settlement in Palestine before 1914 as “colonization,” he also argues that Zionism cannot be described as a “settler-colonial” society according to current prevailing definitions. He defines colonialism as a concept centered around the “control of one people over another,” necessitating the existence of a metropole (an imperial state which exercises power over distant colonies) as well as a relationship between the colonists and the “subject population” (Dowty, 2022). Dowty (2022) makes the point that because Jewish Europeans lacked a true metropole, they can be best defined as persecuted refugees who did not represent the interests of a mother country in their migration to Palestine. Furthermore, he defines settler colonialism as implying “an intention to replace, or even eliminate, the indigenous people and/or culture” (Dowty, 2022). Dowty (2022) argues that Jewish people, in designing the Jewish state, did not seek to “dominate or reshape” the Arab population. Though he makes note of isolated calls for the transfer of Arabs outside areas of Jewish settlement, in contrast to Wolfe (2006) and Khalidi (2020), he states that “if displacement of the Arab population was somehow a concealed element of Zionism, it clearly has been an abject failure” (Dowty, 2022).

Just two years later, Adam Kirsch, a member of the seminar faculty of Columbia University’s Center for American Studies, published his 2024 essay “The False Narrative of Settler Colonialism” in *The Atlantic*. The essay was adapted from his book, *On Settler Colonialism: Ideology, Violence, and Justice*. Kirsch (2024) argues that even in the years before the October 7th attacks occurred, “the Palestinian struggle against Israel had become widely understood by academic and progressive activists as the vanguard of a global battle against settler colonialism,” going on to cite Khalidi’s claim in *The Hundred Years’ War on Palestine* of the ultimate settler-colonial goal of Zionism (Kirsch, 2024). He then states that although Israel “has become the standard reference point” for the contemporary model of settler colonialism, it undoubtedly fails to meet the requirements of such a model in “obvious ways” (Kirsch, 2024).

Similar to Dowty (2022), Kirsch (2024) establishes an operative definition of settler-colonial characteristics and then deconstructs the lack of these characteristics within Israel. Indeed, just as Dowty (2022) concluded in “Is Israel a Settler Colonial State?”, Kirsch (2024) establishes that “the Jewish state did not erase or replace the people already living in Palestine,” though he finds it relevant to note that “it did displace many of them” (Kirsch, 2024). Kirsch (2024) goes further, questioning the concept and use of indigeneity by the Palestinian cause, the associated moral implications, and its role in the discourse surrounding settler colonialism. In stark contrast to Salamanca et al. (2012), Kirsch (2024) argues against the value of comparing Israel-Palestine to other contemporary examples of settler colonialism, such as the ongoing Sinicization of the Tibetan and Uyghur people by the Chinese government. In his view, grouping together causes that are so distinct from one another may actually have the adverse effect of “inhibit[ing] understanding of each individual cause. Any conflict that fails to fit the settler-colonial model must be made to fit” (Kirsch, 2024). Through this lens, the application of the label of settler colonialism distorts an understanding of the Israel-Palestine dynamic more than it allows for constructive comparisons.

Notably, both Dowty (2022) and Kirsch (2024) present Israel’s lack of a traditional metropole as crucial evidence in the argument against viewing Palestine through a settler-colonial lens. Alternatively, historian Tariq Dana provides a rebuttal to this point in particular in his 2024 article “Notes on the ‘Exceptionalism’ of the Israeli Settler-Colonial Project,” published in the *Middle East Critique*. Dana, an Associate Professor and the Head of the Conflict Management and Humanitarian Action program at the Doha Institute for Graduate Studies, concedes that Israel’s form of colonization did in fact lack a conventional metropole, as one might recognize more easily in the colonial projects of the French and British empires. However, he proposes that Israel instead evolved “as a manifestation of Western imperialism with unique, home-grown attributes,” an aberration which he claims plays a central role in the argument for Israel’s exceptionalism regarding settler colonialism (Dana 2024, 167). In the absence of a conventional metropole, Dana suggests that Israel instead developed alliances with Britain and the United States, thereby ensuring continued capital, military, and diplomatic support for the state’s colonial aspirations by “effectively securing a proxy metropole” (Dana 2024, 167). He defines this dynamic as a symbiotic “form of subimperial stewardship,” in which Israel functions within and benefits from the aims of American imperialism as a settler colony (Dana 2024, 167). In “Notes on the ‘Exceptionalism,’” Dana takes the previously established stance of Salamanca et al. (2012) in “Past is Present,” contending that organizing Israel-Palestine within a settler-colonial framework both “illuminates the specific dynamics at play” and “sheds light on the broader implications for understanding and challenging similar structures of oppression worldwide” (Dana 2024, 165).

Looking Toward October 7th, 2023

During the twenty-first century, a notable shift in scholarly discourse surrounding Israel-Palestine and its place in settler-colonial studies has occurred. Many scholars argue that Israel-Palestine is best understood within the context of settler colonialism. Some, building on

the works of Patrick Wolfe, have expanded this argument to include the value of Israel-Palestine as a means of comparison across other historical and contemporary examples of settler colonialism. On the other hand, counterarguments have also emerged in substantial numbers, emphasizing the reductive nature of the settler-colonial label; frequently cited evidence includes Israel's lack of a true metropole, as well as the extensive historical Jewish connection to the land of Israel.

The attacks of October 7th, 2023, undoubtedly thrust the issue of Israel and Palestine to the forefront of the global stage, due in no small part to the scale of the attacks in combination with the extensive media coverage. In 2024, New York-based reporter Jennifer Schuessler published the article, "What is 'Settler Colonialism'?" in the *New York Times*, attempting to provide a general outline of the history of the term "settler colonialism," as well as its usage in numerous cultural movements and scholarly publications since the twentieth century. She devotes a section of the article to discussing in greater detail the controversy surrounding Israel and the settler-colonial label, stating that "nowhere is the idea of settler colonialism more charged than in discussions of Israel, whether it is used to describe Israel's current settlements in the West Bank or the processes that led to the founding of the Jewish state itself in 1948" (Schuessler, 2024).

Schuessler cites numerous scholars on the subject, ranging from Rashid Khalidi's *The Hundred Years' War* to Tom Segev's *One Palestine, Complete: Jews and Arabs Under the British Mandate*; in each case, a different justification is presented as to whether or not Israel-Palestine is best understood through a settler-colonial lens. Some, such as Khalidi, have made clear they believe the settler-colonial paradigm is necessary to build a more comprehensive understanding of Israel-Palestine. Others, like Segev, have argued that "colonialism is irrelevant to the Zionist experience" (qtd. in Schuessler, 2024). Still others, including political scientist Aziz Rana, have taken a slightly different approach, allowing that the settler-colonial paradigm "doesn't tell you everything you need to know," while maintaining that "it allows you to see something that you otherwise would not be able to see." (qtd. in Schuessler, 2024).

Of course, this article is merely a snapshot of the massive upsurge of discussion surrounding the topic, and the polarization of opinions that has only continued to intensify amidst the continued War in Gaza. Since the October 7th attacks and Israel's subsequent retaliatory military action, coverage on Palestine and the War in Gaza by news, social media, and international political discussion has increased exponentially, propelling the longstanding discussion of Israel-Palestine into even greater prominence. The subject of settler colonialism as it pertains to Israel and Palestine is indisputably a landmine, now more so than ever, as settler-colonial studies gains more traction in the academic world. Nevertheless, in coming to an understanding of Israel-Palestine and debating the legitimacy of the settler-colonial paradigm as applied to Palestine studies, it is crucial that decades of history and scholarly discussion of settler colonialism are taken into account, as opposed to the flattening and sidestepping of valuable historical context that has become all too common.

Works Cited

- Dowty, Alan. "Is Israel a Settler Colonial State?" University of Washington Stroum Center for Jewish Studies, 10 Nov. 2022, jewishstudies.washington.edu/israel-hebrew/why-israel-isnt-a-settler-colonial-state/.
- Gershon Shafir, "Zionism and Colonialism: A Comparative Approach", in Ilan Pappé, ed., *The Israel/Palestine Question* (London: Routledge, 1999), 81–96.
- Hamzah Baig, "Spirit in Opposition: Malcolm X and the Question of Palestine," *Social Text* 140 37, no. 3 (2019): 47-71.
- Khalidi, Rashid. *The Hundred Years' War on Palestine: a History of Settler Colonialism and Resistance, 1917-2017*. Metropolitan Books, Henry Holt and Company, 2020.
- Kirsch, Adam. "The False Narrative of Settler Colonialism." *The Atlantic*, 20 Aug. 2024, www.theatlantic.com/ideas/archive/2024/08/how-settler-colonialism-colonized-universities/679514/.
- Lorenzo Veracini (2013) 'Settler Colonialism': Career of a Concept, *The Journal of Imperial and Commonwealth History*, 41:2, 313-333, DOI: 10.1080/03086534.2013.768099
- Omar Jabary Salamanca, Mezna Qato, Kareem Rabie & Sobhi Samour (2012) Past is Present: Settler Colonialism in Palestine, *Settler Colonial Studies*, 2:1, 1-8, DOI: 10.1080/2201473X.2012.10648823
- Patrick Wolfe (2006) Settler colonialism and the elimination of the native, *Journal of Genocide Research*, 8:4, 387-409, DOI: 10.1080/14623520601056240
- Piterberg, Gabriel. "Israeli Sociology's Young Hegelian." *Journal of Palestine Studies*, vol. 44, no. 3, 2015, pp. 17-38. JSTOR, <https://doi.org/10.1525/jps.2015.44.3.17>. Accessed 8 Sept. 2025.
- Schuessler, Jennifer. "What Is 'Settler Colonialism'?" *The New York Times*, 22 Jan. 2024, www.nytimes.com/2024/01/22/arts/what-is-settler-colonialism.html.
- Tariq Dana (2024) Notes on the 'Exceptionalism' of the Israeli Settler-Colonial Project, *Middle East Critique*, 33:2, 165-172, DOI: 10.1080/19436149.2024.2342733

A Comprehensive Review of Artificial Intelligence Applications in the Stock Market

By Lindsey Mills

Abstract

This research paper provides a literature review of artificial intelligence (AI) applications in the stock market, with a focus on forecasting stock market volatility, risk management, and portfolio optimization. This paper consolidates existing research on the key AI models Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL) to evaluate its implications in the stock market, benefits, drawbacks, and future research questions. Findings show that the use of NLP enhances forecasting stock market volatility by evaluating investor and media behavior, while ML is also able to accurately predict stock price movements. The application of NLP and ML have demonstrated positive contributions to risk management by analyzing large volumes of numerical, textual, and auditory data. Finally, ML, either independently or combined with NLP, DL, and RL, has improved the accuracy of determining asset return and risk for portfolio optimization. The limitations of the use of these AI models consist of potential algorithmic bias, lack of model transparency, ethical concerns, unequal access to AI models, and inaccuracy.

1. Introduction

Artificial intelligence (AI) is a technology that uses computer and machine technology to process data and automate tasks; it has been a transformative tool incorporated into many industries. For instance, over recent years there has been a significant rise in the use of AI in the stock market. AI has primarily been used for more accurate and efficient stock market volatility forecasting, risk management, and portfolio optimization. The main AI models used in the stock market are Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL).

NLP is an AI model that focuses on processing human language, such as written texts like tweets, and in the context of the stock market, it can be used to determine public sentiment toward a specific stock (Mehtab 2019). When incorporated into the stock market, the main use of this AI model is to perform sentiment analysis: the act of determining the opinion or tone of a certain piece of media. This is useful for the stock market because if the AI can calculate the overall mood toward a certain company, investors will know whether to avoid a company or not.

Machine Learning (ML), Deep Learning (DL), or a combination of the two, have also been incorporated into the stock market. These versatile models are mainly used for stock price prediction, risk management, and portfolio optimization (El Hajj 2023; Aziz 2019; Faheem 2022). Machine Learning refers to algorithms that are able to find patterns in a data set and learn from its past predictions to be more and more accurate every time (Aziz 2019). Deep Learning is a subsection of Machine Learning; however, it is more programmed to its specific task and mimics the learning process of the brain to predict a certain value (El Hajj 2023). This method of using Machine Learning and Deep Learning artificial intelligence is able to predict the future by

using past data to establish a likely pattern that the stock value will follow. Reinforcement Learning (RL) is another subsection of Machine Learning and its main application has been for the task of portfolio optimization (Sutiene 2024). RL differs from ML because the AI model learns from testing out different decisions to see how good or bad they end up based on the positive outcome; in other words, while ML determines patterns, RL finds the best rewards over time (Meng 2019).

The purpose of this review paper is to clearly outline the benefits, drawbacks, and future research directions of AI in the stock market, specifically for the purpose of stock market volatility forecasting, risk management, and portfolio optimization. This paper contributes to AI literature by synthesizing existing research to evaluate the advantages, disadvantages, and future research questions of these three sections, which are summarized in a table. This paper is meant to inspire future research on this topic and encourage responsible AI development in the financial field.

2. Method

A literature review was conducted by collecting research in the fields of artificial intelligence, finance, and the stock market. The keywords “Artificial Intelligence,” “Machine learning,” “Natural Language Processing,” “Deep Learning,” “Reinforcement Learning,” “Stock market,” “Forecasting stock market volatility,” “Risk Management,” and “Portfolio Optimization” were used to collect research in the fields of AI and finance on Google Scholar, Science Direct, and ResearchGate.

3. Results

3.1 The Application of AI on Forecasting Stock Market Volatility

Stock market volatility is defined as the variation in stock prices, which is determined by calculating the magnitude of a change in stock price and the rate at which it was made (Pataci 2022). Research has shown that advancements in Natural Language Processing and Machine/Deep Learning technologies have improved their ability to accurately predict future stock market trends using sentiment analysis and stock price prediction. Utilizing AI to process data such as news reports, social media, and past market data can lead to more informed investments strategies.

NLP in Sentiment Analysis

Sentiment Analysis in the stock market determines the overall mood toward a certain stock or company that can be invested in (Sahini 2024). Using Natural Language Processing (NLP), Artificial Intelligence is important for determining stock market volatility because, although its goal is not to predict an actual stock price, it can be used to calculate the tone of media sources toward the company to find if it is generally positive or negative. Essentially, traditional financial methods are unable to fully consider the impact of human bias on stock

market trends. Media sources like social media have also made the data for sentiment analysis more abundant. For sentiment analysis, Dupont and El-Sayed (2024) show that Sentiment Analysis tools elevate the capability of forecasting stock market volatility by analyzing emotional and psychological factors that influence investor behavior. Calculating the overall mood of investor behavior is not something that other AI can perform; so, NLP is a unique form of AI that can utilize textual and speech data for its task. Also, a study conducted by Puh and Bagic Babac (2023) concluded that there is a relation between the conclusions of NLP from news headlines and the predicted stock values. This discovery reveals that this kind of AI, while not predicting an actual value, can predict the performance of the stock market through other means. On the other hand, Dupont and El-Sayed (2024) explain that the use of NLP for stock market volatility is not entirely reliable because the lack of transparency in the model suggests that there could be “market manipulation.” This means that an excess of biased information can be released on informational platforms such as social media which would affect the conclusion of the AI.

Deep Learning and Machine Learning in Stock Price Prediction

Machine Learning (ML) and Deep Learning (DL) are vital in predicting actual stock market values based on historical data and established patterns. Dau et al. (2024) found that ML helps with the accuracy of stock market predictions. Also, DL and ML can work together for improved accuracy. This combination not only allows for the predictions of data patterns, but it is also able to adapt to more nuanced market behaviors. Moreover, ML and DL are able to be used separately and together for accurate results of stock price prediction. Dau et al. (2024) show that DL has allowed for the prediction of stock prices to a significantly accurate degree. However, a shift from ML techniques to DL techniques has been observed, demonstrating the constant evolution of the financial world (Dau 2024).

Overall, sentiment analysis and stock market price prediction would not be as efficient and accurate without the use of the AI models. A survey conducted by Mane (2022) proved that over time the advancements in the AI technologies have improved their ability to accurately predict future stock market data. However, it is necessary that the data remains unbiased and ranges widely for the most accurate depiction of future trends.

3.2 The Application of AI for Risk Management

Risk Management is defined as the investor’s ability to balance their potential gains and losses in the stock market (Divya 2019). The process of risk management, as stated by Darwiesh et al. (2024), is a cycle of identifying risks, assessing risks, handling risks, and monitoring risks. So, the usage of AI has become a significant contributor to managing the risks of investments efficiently. According to Agarwal et al. (2024) the combination of many different AI tools has had a compounding positive effect to completing risk management in the financial market. This is because AI is efficient at predictions, and it also has a very fast operation speed. Although there are many different AI models that are used for risk management, the two most common

models are Natural Language Processing for Sentiment Analysis and Machine Learning for data analysis.

Sentiment Analysis Using NLP for Risk Management

The use of NLP for risk management is becoming a very common use of AI in the stock market. Over time, the NLP models are improving in complexity and in their ability to process large amounts of written and audio data to reach a conclusion. According to Sahani (2024), it is expected that AI in the stock market will become even more refined, especially for NLP. So, the “predictive power” of not only NLP, but all AI models, will improve. Additionally, social media has created a large amount of data for the use of NLP since the AI is able to analyze the behavior of consumers and investors. The findings of Darwiesh et al. (2024) have shown that this has been invaluable to the risk management sector for the stock market. This is because the AI is able to process a large amount of data to then identify trends and patterns which allows investors to make decisions that are data driven (Kuzior 2024).

Sentiment analysis using NLP for Risk Management is an important tool to limit risks of loss in the stock market. However, there are downfalls to the use of this AI. For instance, although Sentiment Analysis helps investors make relevant data-driven decisions, there are still challenges such as algorithmic bias, lack of data privacy, and the unpredictability of human emotions (Sahani 2024). Social media can also be a downfall because there is a large amount of data that could lead to an inaccurate answer (Sahani 2024).

Machine Learning for Risk Management

The use of Machine Learning for risk management in the financial sphere has become a very promising resource. Rahmani et al. (2020) show that Machine Learning can be used for risk management to analyze financial data for risks such as credit risk, market risk, and operational risk. Essentially, Machine Learning is able to predict future trends which allows for investors to make more data-driven decisions. So, the potential for an investment in a stock that results in a net loss is much lower. Kuzior et al. (2024) also has an optimistic view toward ML for Risk Management. For instance, Kuzior et al. (2024) mention that being able to identify and mitigate risks is directly linked to a more stable financial system.

While ML is a useful AI model for risk management, there are risks associated with using it. The use of Machine Learning and AI in financial markets has grown over recent years for various financial operations including risk management. However, there is a need for surveillance over data privacy (El Hajj 2023). As AI continues to grow, it is possible that privacy and ethical issues such as biased decisions or using unauthorized data may not be controlled. A loss of data privacy could result in data leaks and investor mistrust. Additionally, El Hajj (2023) found that there is a possibility of job replacement from the incorporation of this AI into financial companies. Finally, Aziz (2019) found that the management of data collection and transparency of the AI models can also be a limitation. This is because the data must be reliable and the investors should be able to trust the outputs of the AI. Overall, Machine Learning is a

highly utilized AI tool for operations in the stock market, including risk management. However, there are still risks and limitations that should be considered as it continues to grow in complexity and popularity.

3.3 The Application of AI on Portfolio Optimization

Portfolio Optimization is the process of selecting investment portfolios that have an optimal balance between risk and return for investment (Santos 2022). Over recent years, AI models, primarily ML, have been incorporated into the financial market to assist with this task. Before the increase in the use of AI for Portfolio Optimization, more traditional and less automated models were used. For instance, Faheem (2022) notes that the original strategy for optimizing an investment portfolio was the Markowitz model. The Markowitz model used the strategy of balancing the factors of return and risk using a “diversification principle.” However, this came with the issue of a limited ability to adapt to a constantly changing market with a high volatility (Faheem 2022). Essentially, AI has been able to expedite the process of portfolio optimization and do so with accuracy.

ML for Portfolio Optimization

According to Faheem et al. (2022), ML has transformed the financial field by optimizing and managing portfolios in order to make improved financial decisions about the collection of investments. Sutiene et al. (2024) found that the use of ML improves asset return and risk estimates of portfolios. Additionally, Sutiene et al. (2024) argue that the cost of information is lower with the introduction of AI for all portfolio managers. This potential for universal access to portfolio optimizing technology could lead to a more efficient way to access information for portfolio management and create a more stable financial market. As AI continues to improve, Faheem et al. (2022) predict that developments of ML will further improve its ability to optimize portfolios. Overall, Artificial Intelligence using ML techniques has improved the efficiency and accuracy for the portfolio optimization process by making it more automated.

The process of portfolio optimization has generally improved using AI; however, there are improvements that can be made to support the potential downfalls of using the technology. Faheem et al. (2022) and Sutiene et al. (2024) observe that the conclusions made by the AI should be more interpretable to allow financial investors to trust the AI and its conclusions. Finally, recent research shows that the potential unequal access to AI would create an unequal opportunity for portfolio managers and other financial actors in the stock market (Sutiene 2024). Thus, the transparency and accessibility of the AI must be improved in order for the technology to be a more effective tool in the stock market.

ML Combined with DL, RL, or NLP for Portfolio Optimization

Although ML has been the main AI model used for Portfolio Optimization, similar AI models such as Deep Learning (DL), Reinforcement Learning (RL), and NLP have been combined with ML for portfolio optimization (Santos 2022; Sutiene 2024). Santos et al. (2022)

found that hybrid models, which include the use of two or more algorithms at once, are being used to improve the results of the AI to construct a portfolio. As a result, portfolio optimization can become more advanced by combining different AI models due to their varied capabilities in data analysis and pattern recognition in a highly volatile stock market. Also, Sutiene et al. (2024) found that using NLP has been very helpful to manage portfolios that will maximize the relationship between risk and return.

4. Discussion

Overall, the use of AI has significantly improved the efficiency and accuracy of forecasting stock market volatility, risk management, and portfolio optimization in the stock market. The AI models including ML, NLP, DL, and RL are the primary contributors to the shift toward a more data driven, computational, and stable stock market.

Although many advantages to using AI in the stock market have been identified, there are still downfalls that may inhibit its continued growth, which provides direction for future research. For instance, one of the major concerns across the three main uses of AI discussed earlier is that AI can be untrustworthy. So, model transparency must improve since it is valuable to know how the data is sourced and what the underlying process is for the AI to provide its output. Also, issues of data privacy and ethical concerns when using AI still remain. Finally, since the stock market is unpredictable, it is necessary to understand the limitations to using AI for these relevant tasks.

In the future, AI will continue to improve and become more advanced, which could potentially lead the way for new strategies and uses of AI to be incorporated into the financial market. Similar to how a combination of ML, DL, and RL is used for portfolio optimization (Santos 2022), it is possible that more hybrid uses of AI will be used in many more ways in the stock market. Also, the improvements of the AI models will address and minimize the current limitations of using AI in the stock market. Table 1 summarizes the advantages, challenges, and future research questions for the three main uses of AI in the stock market.

Table 1: A summary of AI's advantages, challenges, and future research questions

Use of AI	Advantages	Challenges	Future Research Questions
Forecasting Stock Market Volatility	Natural Language Processing - Can process a wide range of data (social media, news articles, etc) Machine/Deep Learning - Accurate stock market price prediction	- Inaccuracy, data quality, and computational complexity - Limited model transparency - False/filtered data can lead to inaccurate predictions of mood toward stock	- How will the issue of biased posts on the internet (social media or news sources) be resolved to ensure a more accurate depiction of the mood toward a certain stock? - As AI improves, will the technology be

			able to process more types of data like photos?
Risk Management	Natural Language Processing • Increased computational complexity • Can process large pools of data Machine Learning • Fast-growing and efficient AI model for predicting trends	• Unpredictability of human emotions • Data privacy and ethical concerns • Possible job replacement • More model transparency is needed	• Will it be possible to incorporate AI into jobs in the stock market to prevent job replacement? • How can ML and NLP be improved or combined with other AI models to improve its accuracy?
Portfolio Optimization	• Potential combination of AI models	• Limited interpretability • Unequal access for portfolio managers	• Could the issue of limited interpretability be solved for the AI models? • How can the issue of unequal access to AI in the stock market be prevented?

5. Conclusion

In conclusion, recent research has shown an optimistic view of the incorporation of AI into the stock market for the processes of forecasting stock market volatility, risk management, and portfolio optimization. AI models such as NLP, ML, DL, and RL in the finance sphere have become a useful tool due to its strengths of efficient data analysis, accurate predictions, and pattern recognition. However, research has also shown concerns such as lack of computational complexity, limited model transparency, and ethical consideration. Also, as it was observed by Sutiene (2024), the possibility of unequal access to AI will create unfair advantages in the stock market. Thus, it is important that the spheres of finance and technology continue to innovate the AI models to solve these issues in order to avoid any complications in the future. Overall, while AI has the potential to be a transformative tool in the financial field, it is necessary that it is developed responsibly to ensure its long-term success.

Works Cited

- Agarwal, Mukesh, et al. "AI-Driven Risk Management in Financial Markets and Fintech." 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES). (2024)
- Aziz, Saqib, and Michael Dowling. "Machine Learning and AI for Risk Management." *Disrupting Finance* (2019): 33-50.
- Dao, Hong N., et al. "Ai in Stock Market Forecasting: A Bibliometric Analysis." *SHS Web of Conferences*. Vol. 194. EDP Sciences, (2024)
- Darwiesh, Abdelaziz, A. H. El-Baz, and Mohamed Elhoseny. "Intelligent Risk Management System for Enhancing Performance of Stock Market Applications." *Expert Systems with Applications* 249 (2024): 123493.
- Divya, T. S., and A. M. Viswambharan. "Investment Risk Management." *Shanlax International Journal of Commerce* 7.4 (2019): 36-41.
- Dupont, Nicolas, and Ahmed El-Sayed. "A Review of Sentiment Analysis Tools in Predicting Stock Market Volatility: AI and Text Mining Approaches." *Business, Marketing, and Finance Open* 1.2 (2024): 41-52.
- El Hajj, Mohammad, and Jamil Hammoud. "Unveiling the Influence of Artificial Intelligence and Machine Learning on Financial Markets: A Comprehensive Analysis of AI Applications in Trading, Risk Management, and Financial Operations." *Journal of Risk and Financial Management* 16.10 (2023): 434.
- Faheem, Muhammad, Muhammad Aslam, and Sridevi Kakolu. "Artificial Intelligence in Investment Portfolio Optimization: A Comparative Study of Machine Learning Algorithms." *International Journal of Science and Research Archive* 6.1 (2022): 335-342.
- Ferreira, Fernando GDC, Amir H. Gandomi, and Rodrigo TN Cardoso. "Artificial Intelligence Applied to Stock Market Trading: A Review." *IEEE Access* 9 (2021): 30898-30917.
- Kuzior, Aleksandra, Artem KOLDOVSKIY, and Ihor REKUNENKO. "Optimizing Financial Market Stability through AI-based Risk Management." *Materials Research Proceedings* 45 (2024).
- Li, Audeliano Wolian, and Guilherme Sousa Bastos. "Stock Market Forecasting using Deep Learning and Technical Analysis: A Systematic Review." *IEEE Access* 8 (2020): 185232-185242.
- Mane, Om. "Stock Market Prediction using Natural Language Processing--A Survey." *arXiv preprint arXiv:2208.13564* (2022).
- Mehtab, Sidra, and Jaydip Sen. "A Robust Predictive Model for Stock Price Prediction using Deep Learning and Natural Language Processing." *arXiv preprint arXiv:1912.07700* (2019).
- Meng, Terry Lingze, and Matloob Khushi. "Reinforcement Learning in Financial Markets." *Data* 4.3 (2019): 110.
- Pataci, Hilal, et al. "Stock Price Volatility Prediction: A Case Study with AutoML." *Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing*

- (FinNLP). (2022)
- Puh, Karlo, and Marina Bagić Babac. "Predicting Stock Market using Natural Language Processing." *American Journal of Business* 38.2 (2023): 41-61.
- Sahani, Tanwangini. "Decoding Market Emotions: The Synergy of Sentiment Analysis and AI in Stock Market Predictions." *Journal of Next-Generation Research* 5.0 (2024).
- Santos, Gustavo Carvalho, et al. "Portfolio Optimization using Artificial Intelligence: A Systematic Literature Review." *Exacta* 22.3 (2024): 766-787.
- Sutiene, Kristina, et al. "Enhancing Portfolio Management using Artificial Intelligence: Literature Review." *Frontiers in artificial intelligence* 7 (2024): 1371502.
- Rahmani, Amir Masoud, et al. "Applications of Artificial Intelligence in the Economy, Including Applications in Stock Trading, Market Analysis, and Risk Management." *IEEE Access* 11 (2023): 80769-80793.

FinCoach: A Lightweight AI-Powered Financial Coaching System for Students

By Minh Duc Dang Vu

Abstract

This paper presents FinCoach, an AI-powered financial coaching system designed to assist students in developing healthier financial habits. The system combines a classification framework—categorizing students into fifteen financial behavior types based on saving and spending patterns—with a prompt-driven advisory mechanism that generates personalized strategies, budget allocations, and financial warnings. By merging rule-based classification with lightweight language models optimized for low-resource devices, FinCoach provides a scalable, accessible, and adaptive financial guidance tool. A pilot deployment with a small group of students over a four-week period demonstrated preliminary improvements in saving consistency and reduced impulsive spending, highlighting FinCoach’s potential to promote sustainable financial literacy among students while addressing the limitations of existing budgeting applications.

Introduction

One of the most common challenges faced by students today is the difficulty of managing personal finances effectively. Research indicates that young adults often struggle to balance short-term consumption with long-term savings, resulting in financial stress and poor decision-making (Taylor et al., 2025; Ng et al., 2020). Existing digital tools and finance apps provide general-purpose tracking but are rarely tailored to the irregular income patterns and limited financial literacy levels of students.

This paper contributes to addressing this gap by (1) introducing a fifteen-type financial behavior classification framework, (2) integrating lightweight AI models for efficient advisory generation, and (3) presenting a pilot study of FinCoach, an AI-driven chatbot that combines financial behavior analysis with real-time spending feedback to support students in building healthier financial habits.

Design

This study adopts an applied AI approach using a pilot intervention, allowing for iterative adjustments and refinements based on user feedback and observed outcomes. The intervention consists of deploying the FinCoach chatbot to a sample of students over a defined period and monitoring changes in their financial behavior.

The FinCoach system was designed with three main components.

1. User Input Layer

Students provide financial information through surveys or optional spending data uploads. The system gathers both behavioral (saving and spending tendencies) and contextual (financial situation) inputs.

2. Classification Engine

Inputs are mapped onto a structured framework of fifteen financial behavior types. Each type is associated with predefined risk factors, spending patterns, and tailored strategies.

3. Advisory Generator

Based on classification results, the system creates customized recommendations, including saving targets, category-based budget allocations, warnings, and long-term strategies. Outputs are displayed in a clear and student-friendly format, with updates provided on a weekly and monthly basis.

This modular design ensures that FinCoach remains interpretable, scalable, and adaptable to different student populations.

Models

To power the advisory generator, FinCoach relies on compact and efficient natural language models optimized for low-resource environments. Key consideration includes:

Category	GPT-OSS-20B	RWKV7-G1 “GooseOne”	Google Flan-T5 Small / Base	Qwen 2 7B Instruct
Architecture	Transformer (decoder-only)	Pure RNN (state-based, transformer-free)	Encoder–decoder transformer	Transformer (decoder-only, instruction-tuned)
Parameters	~20B	~2.9B	~80M (small) ~250M (base)	~7B
Reasoning Style	Pattern-based reasoning, strong generative capacity	Sequential reasoning, memory-efficient long context	Instruction-tuned, strong zero/few-shot reasoning	Balanced reasoning + fluency; optimized for dialogue
Speed / Memory Efficiency	Needs ≥40GB GPU VRAM; slow inference	Very efficient on CPU/RAM; minimal VRAM	Efficient relative to size; scales across hardware	Runs on single modern GPU; supports 4-bit quantization
Strengths	High-quality generation, broad knowledge base	Lightweight, efficient, strong reasoning for its size	Excellent instruction-following and reasoning	Strong reasoning + conversation, multilingual support
Weaknesses	Heavy compute requirements; less practical for personal use	Less fluent for open-domain creative text	Less natural for long, open-ended generation	Needs GPU; heavier than Flan-T5 or RWKV
Best Use Cases	Research, creative writing, text generation	On-device assistants, reasoning-heavy chatbots	Education, structured QA, reasoning tasks	General-purpose chatbots, reasoning, student/assistant tools

Table 1: Different types of models

Although GPT-OSS-20B exemplifies the scale and generative capacity of modern transformer architectures, its practical utility on resource-limited devices is negligible. Running such a model locally on an 8 GB consumer machine would be tantamount to “fitting a supercomputer into a pocket calculator”—a technically fascinating, but ultimately futile, endeavor. In contrast, RWKV7-G1 “GooseOne” demonstrates that architectural efficiency can outpace sheer parameter count, offering viable reasoning performance within the constraints of modest hardware. Flan-T5, while far smaller, excels in instruction-following tasks but lacks the fluency and depth of larger transformer models. Qwen-2-7B Instruct strikes a middle ground, offering strong reasoning and dialogue abilities, though it remains too heavy for most low-end devices without quantization.

It should also be noted that many users can appear to run GPT-OSS-20B seamlessly on low-end laptops only because Hugging Face provides hosted inference for this model. In such cases, the computation is performed entirely on Hugging Face’s servers, and the local machine merely sends and receives API calls. Thus, the illusion of deploying a 20-billion-parameter model on 8 GB hardware is in fact the by-product of cloud infrastructure rather than local capability.

Procedure

The system operates in three stages:

1. Classification

Students complete a financial behavior survey and may optionally provide actual spending data. Based on survey responses, students are classified into one of fifteen financial personality types according to both their income level and spending tendencies.

Rich - Average - Struggling

Very impulsive - Impulsive - Balanced - Thrifty - Very thrifty

Group	Expenditure Food/Entertainment/Education/Saving	Assessment	Strategy
Very impulsive – Rich	35 / 30 / 20 / 15	Good income but spends recklessly, risk of forming bad habits	Set automatic spending limits; prioritize small investments, avoid emotional spending
Impulsive – Rich	30 / 25 / 25 / 20	Likely to overspend without close monitoring	Use an expense-tracking app; split funds right after receiving salary

Balanced – Rich	30 / 20 / 30 / 20	Spending is stable, but financial opportunities may be missed	Start saving for long-term goals; categorize investment types
Thrifty – Rich	25 / 15 / 35 / 25	Manages money well, could raise financial goals	Try investing / create an education fund; consider allocation for future goals
Very thrifty – Average	20 / 10 / 40 / 30	May save excessively, sacrificing life experiences	Allow yourself to “enjoy” reasonably; move part of savings into asset accumulation
Very impulsive – Average	40 / 30 / 20 / 10	Often runs out of money mid-month, uncontrolled spending	Divide money weekly; set alerts when spending exceeds 75% of budget
Impulsive – Average	35 / 25 / 25 / 15	Needs to build expense-tracking habits	Record daily expenses; allocate clearly by category
Balanced – Average	30 / 20 / 30 / 20	Stable but lacks clear financial strategy	Set specific saving goals; try a “weekly saving fund”
Thrifty – Average	25 / 15 / 35 / 25	Only risk is inconsistent saving	Fix a specific day to transfer into savings; increase emergency fund
Very thrifty – Average	20 / 10 / 40 / 30	Needs balance between saving and life experience	Keep saving consistently, increase spending for self-development if possible
Very impulsive – Struggling	40 / 35 / 15 / 10	Easily falls into debt or has to ask for more money	Strongly recommend dividing money daily; limit entertainment spending
Impulsive – Struggling	35 / 30 / 20 / 15	No planning, lacks emergency funds	Prioritize essential spending, build expense-recording habits, create a small emergency fund
Balanced – Struggling	30 / 20 / 30 / 20	Fine if disciplined, but may run short at month’s end	Limit entertainment weekly, create “envelope” system for partial spending

Thrifty – Struggling	25 / 15 / 35 / 25	Doing well, but needs more tracking tools	Use chatbot for weekly check-ins; automate fund splitting at start of month
Very thrifty – Struggling	20 / 10 / 40 / 30	Very strong control, but risk of stress or lack of experiences	Maintain light saving, prioritize education, consider increasing income via freelancing

Table 2: Classification groups for FinCoach

2. Personalized Advisory Generation

Building on the initial classification, FinCoach produces individualized outputs tailored to each financial personality type. Students are provided with monthly saving goals expressed as percentages of their income, alongside a detailed budget allocation across food, education, entertainment, and reserve funds. To address behavioral risks, the system also delivers financial warnings related to impulsivity, overspending, or excessive restriction, and offers strategies aimed at promoting sustainable long-term financial growth.

3. Feedback and Monitoring

FinCoach does not stop at planning but actively monitors user behavior over time. Students receive weekly and monthly evaluations of their spending performance, while automated alerts are issued if budget thresholds are exceeded. Recommendations adapt dynamically to observed financial patterns, ensuring that users remain aligned with their saving targets and gradually develop healthier financial habits.

Prompt Design and Objective

The effectiveness of FinCoach relies not only on the underlying model but also on the careful design of prompts that guide the chatbot’s reasoning. Each prompt serves to direct the model toward producing structured, actionable financial advice rather than generic text. The goal of the answer is always made explicit in the prompt itself. For example, when the objective is classification, the prompt specifies:

“Classify the user into one of the 15 financial personality types based on their reported income and spending habits. Clearly state the assigned type, explain the behavioral tendencies, and identify both risks and strengths.”

When the objective is planning, the prompt makes this explicit:

“Based on the user’s financial type and goals, generate a weekly or monthly budget plan broken into clear categories (Food, Entertainment, Education, Savings). Ensure the output is concise, practical, and student-friendly.”

This approach ensures that the model not only produces an answer but does so in a way aligned with the chatbot's coaching objectives. By embedding the desired structure and focus directly in the prompt, FinCoach reduces variability in outputs and increases the likelihood of consistent, useful responses for students.

→ Final prompt:

You are FinCoach, a personal money coach for students. Always follow these rules:

1. Classification

Classify the user into one of 15 financial types based on spending habits and financial situation.

Clearly explain their habits, strengths, and risks.

2. Planning

If the user asks for a plan, provide a weekly or monthly breakdown that is:

Practical and student-friendly

Written as clear bullet points

Organized by days (Mon–Sun) or categories (Food, Education, Entertainment, Savings, etc.)

Keep plans realistic and focused on healthy habits.

3. Data Analysis

When the user provides income and spending data, calculate and explain:

Spending speed, saving rate, non-essential ratio, distribution (first 2 weeks vs last 2), stability.

Interpret as healthy, risky, or needs adjustment.

4. Personalized Recommendations

Always provide:

A monthly saving goal

A budget allocation

Warnings and strategies

5. Style

Direct, concise, positive but realistic

Use everyday language

Final product: FinCoach_Final

The final product of FinCoach integrates a comprehensive set of features designed to support students in developing healthy financial habits. At its core, the system begins with classification, where users are categorized into one of 15 financial types based on their saving and spending patterns as well as their financial situation. Once classified, the chatbot provides personalized financial guidance, including monthly saving goals (expressed as percentages),

tailored budget allocations across categories like food, education, and entertainment, risk warnings, and recommended strategies. It incorporates persistent chat history to ensure continuity, allowing the system to track past spending behaviors, highlight recurring issues, and provide meaningful follow-ups. The platform also includes budget tracking tools that allow users to log actual spending, receive weekly budget allocations, and get real-time alerts when limits are exceeded. At the end of each cycle, monthly evaluations summarize performance, compare spending against goals, and adjust recommendations. Additionally, FinCoach offers weekly and monthly feedback reports, suggesting corrective actions and new saving goals to keep users on track. Together, these functions make FinCoach an intelligent, adaptive, and supportive financial coach tailored for student needs.

Example User Interaction

After Classification:

 Name: Minh Duc Dang Vu

 DOB: 04/07/2008

 Income: \$135 / month

 Saving Rate: 20%

 Food: 33%,  Fun: 33%,  Education: 14%,  Reserve: 0%

 Financial Type: Thrifty – Rich

 Warning: Good money control, should upgrade financial goals.

 Suggested Strategy: Try investing or learning fund; allocate toward future-oriented goals.

Interaction with chatbot:

User:

What's my name and classification?

FinCoach:

Your name is Minh Duc Dang Vu.

Your financial classification is Thrifty – Rich.

User:

I have 100 dollars for 2 weeks, how should i spend it

Fincoach:

Here's a simple, student-friendly way to spend \$100 over two weeks:

1. Reserve a safety net

Set aside \$20 (20 %) for unexpected costs or a small emergency.

Keep it in a high-interest savings account or a cash-back app.

2. Cover your essentials

Food & groceries – \$30 (30 %)

Plan simple, budget meals (rice, beans, frozen veggies, eggs).

Buy in bulk when possible and use coupons.

3. Keep learning and growing

Education / skill-building – \$15 (15 %)

Allocate to an online course, e-book, or a subscription to a learning platform (e.g., Coursera, Skillshare).

4. Treat yourself

Fun / entertainment – \$15 (15 %)

Use for a movie night, a coffee, or a small outing that feels rewarding.

5. Invest in your future

Future-oriented goal – \$20 (20 %)

Put in a low-fee brokerage account or a robo-advisor, or start a micro-investment app.

Even small, regular contributions build long-term wealth.

Weekly Breakdown (optional)

Week 1: \$50 – split 10 % for safety, 15 % food, 10 % education, 10 % fun, 5 % future.

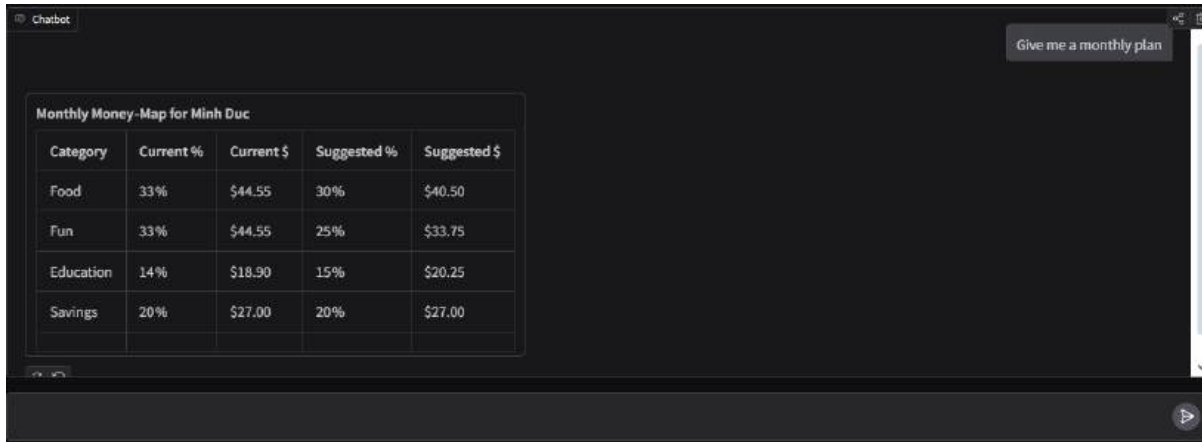
Week 2: Same split – keep the rhythm, adjust if any category needs a tweak.

Stick to the plan, track your receipts, and you'll finish the two weeks with a clear sense of where every dollar went—plus a small savings cushion and a seed for future growth.

Image examples

The image shows a web-based financial classification tool. At the top, there are input fields for 'Full Name' (Minh Duc Dang Vu), 'Day' (01), 'Month' (01), and 'Year' (2000). Below these are four input fields for percentages: '% on Food' (33), '% on Entertainment' (33), '% on Education' (14), and '% Saving' (20). A 'Monthly Income' field shows 135. The 'Classification Result' section displays a summary: Name: Minh Duc Dang Vu, DOB: 01/01/2000, Income: \$135 / month, Saving Rate: 20%, Food: 33%, Fun: 33%, Education: 14%, Reserve: 0%. It also shows the 'Financial Type: Thrifty - Rich', a 'Warning: Good money control, should upgrade financial goals.', and a 'Suggested Strategy: Try investing or learning fund; allocate toward future-oriented goals.' At the bottom, there is a button labeled 'Classify Financial Type'.

Image 1: Example of a manual classification answer



Chatbot

Give me a monthly plan

Monthly Money-Map for Minh Duc

Category	Current %	Current \$	Suggested %	Suggested \$
Food	33%	\$44.55	30%	\$40.50
Fun	33%	\$44.55	25%	\$33.75
Education	14%	\$18.90	15%	\$20.25
Savings	20%	\$27.00	20%	\$27.00

Image 2: Example of an user's question and the chatbot's respond

Future Directions

While the pilot demonstrates promising outcomes, several avenues remain for expansion and refinement:

Outside interface: While the pilot demonstrates promising outcomes, several avenues remain for expansion and refinement, such as enhancing the outside interface to provide a more intuitive, visually engaging, and user-friendly experience.

Chat History: Persistent memory of prior interactions would enable FinCoach to detect patterns and deliver more personalized, context-aware guidance.

Bank Integration: Connecting FinCoach directly to banking apps or e-wallets would allow for automated expense tracking, removing reliance on self-reporting.

Gamification: Introducing challenges, streaks, and reward systems could increase engagement and make financial discipline more enjoyable.

Adaptive Learning: Over time, FinCoach could refine its recommendations by learning from each user's long-term financial behavior.

Expanded Literacy Modules: Beyond budgeting, future versions could incorporate lessons on investing, credit management, and debt avoidance.

Scalability & Deployment: Testing across larger and more diverse student populations would validate FinCoach's effectiveness and adaptability across cultural and economic contexts.

By pursuing these directions, FinCoach could evolve from a pilot chatbot into a comprehensive digital financial mentor, empowering students not only to save and budget effectively but also to build the foundations of lifelong financial literacy.

Conclusion

This pilot study provides preliminary evidence that FinCoach, an AI-powered chatbot, can support students in building healthier financial habits. By combining behavioral classification, tailored budgeting plans, real-time data analysis, and personalized recommendations, FinCoach delivers a structured intervention that extends beyond traditional expense-tracking applications. The integration of lightweight yet capable AI models makes the

system accessible across devices, while prompt-based design ensures outputs remain consistent and student-friendly.

Although findings are based on a limited pilot, results suggest that AI-driven financial coaching holds promise in bridging the gap between awareness and action, enabling students to reduce financial stress and establish habits that contribute to long-term financial well-being.

Works Cited

- Eleuterio DS, Silva J, Pereira A, et al. A Chatbot to Help Promoting Financial Literacy. In: Proceedings of the 14th Symposium on Languages, Applications and Technologies (SLATE). OASICS. 2025;135:7:1–7:13. Available from: <https://drops.dagstuhl.de/storage/01oasics/oasics-vol135-slate2025/OASICS.SLATE.2025.7/OASICS.SLATE.2025.7.pdf>
- Taylor ZW, Marx A, Nixon D, Ray S, Simonds R, Smith M, et al. Artificial Intelligence Chatbots: Are Chatbots Programmed to Provide Financial Wellness Counseling to College Students? HEFWA Research Brief. April 2025. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5238572
- Coach mAIa – AI Financial Coach. MIT Solve. 2024. Available from: <https://solve.mit.edu/solutions/92248>
- Consumer Financial Protection Bureau. Chatbots in consumer finance. CFPB Research Report. 2023. Available from: <https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/>
- Ng M, Pentina I, Sun Y, et al. Simulating the Effects of Social Presence on Trust, Privacy Concerns & Usage Intentions in Automated Bots for Finance. arXiv preprint. 2020. Available from: <https://arxiv.org/abs/2006.15449>
- Salas-Velasco M. Economic and financial education for investment: Evidence from a chatbot intervention. Journal of Economic Education (or related education journal). 2023. Available from: <https://www.tandfonline.com/doi/full/10.1080/00220485.2023.2191594>
- Lakkaraju K, et al. Comparative evaluation of chatbots as financial advisors. arXiv preprint (comparative chatbot study). 2023. Available from: <https://arxiv.org/pdf/2307.07422.pdf>
- Kajiwarra Y. AI literacy for ethical use of chatbots: Will students accept ChatGPT-like tools? Technology & Education article. 2024. Available from: <https://www.sciencedirect.com/science/article/pii/S2666920X24000547>
- Estima A, Peguinho C. Using gamification to promote financial literacy and consumer education among future generations. Proceedings of the 10th International Conference on Higher Education Advances (HEAd'24). 2024. Available from: <https://riunet.upv.es/handle/10251/188990>
- Cannistrà M, et al. The impact of an online game-based financial education tool on students' financial literacy. Journal of Behavioral and Experimental Finance. 2024. doi:10.1016/j.jbef.2024.100XXXX (conference/journal paper — see publisher site)
- Rodríguez K, Saavedra J. Nudging Youth to Develop Savings Habits: Experimental Evidence Using SMS Messages. Working paper / SSRN. 2015. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2650188
- Beshears J, Choi JJ, Laibson D, Madrian BC. Nudging: Progress to date and future directions. Behavioral Public Policy. 2020. Available from: <https://www.cambridge.org/core/journals/behavioural-public-policy/article/nudging-progr>

- ess-to-date-and-future-directions/ (also PMC review:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7946162/>)
- WSAC / MDRC. Improving a Chatbot to Increase Financial Aid Applications: OtterBot evaluation and optimization report. 2025. Available from:
<https://wsac.wa.gov/sites/default/files/2025-Optimizing.Texting.Report.pdf>
- Zhu H, Li X, Chen J, et al. Implementing artificial intelligence-empowered financial advisory services: design and challenges. *Journal of Service Research / Industry review*. 2024. Available from: <https://www.sciencedirect.com/> (publisher page)
- “Can AI Help with Your Personal Finances?” arXiv preprint. 2025. Available from:
<https://arxiv.org/abs/2412.19784>
- Immediate. Immediate launches AI Financial Coaching. Company announcement / applied product case. 2025. Available from:
<https://www.joinimmediate.com/post/ai-financial-coaching>
- MyFinance AI / Talasila SD. AI-Driven Personal Finance Management: Revolutionizing Budgeting and Financial Planning (pilot study). 2024. Available from:
https://www.researchgate.net/publication/382679575_AI-Driven_Personal_Finance_Management_Revolutionizing_Budgeting_and_Financial_Planning
- Labadee L, et al. Role of AI chatbots in education: systematic literature review. *Educational Technology Research & Development*. 2023. Available from:
<https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-023-00426-1>
- R. Dur, et al. A social norm nudge to save more: Field experiment at a retail bank. *Journal of Public Economics*. 2021;198:104–116. doi:10.1016/j.jpubeco.2021.104348
- CFPB / Policy reviews and academic analyses about AI in finance and consumer protection (e.g., reports on risks, transparency, and consumer outcomes). 2023–2025. Available from:
<https://www.consumerfinance.gov/>

Crisis Response, Pricing Decisions and Global Trends: A Micro and Macro-Economic Study of the Food and Beverage Industry By Vivan Agarwal

Abstract

This paper analyzes the interface of microeconomics and macroeconomics in the food and beverages sector based on three core themes. First, it looks into the business responses to shocks, especially the pandemic, and the way the strategies for resilience were conceptualized at the firm and the sectoral levels. Secondly, it analyzes the menu constitution and price setting based on economic fundamentals such as cost structure and consumer behavior and elasticity. Thirdly, it looks at how domestic businesses are impacted by the larger macroeconomic forces such as global trends in inflation, chain reactions in the supply chain and changes in international demand. The paper concludes that microeconomic elements define the survivability and the operational adaptations of individual firms while the macroeconomic forces define the space for the operations within which the firms move. Locating the food and beverages sector in this space between micro and macroeconomics, then, highlights how the sector reflects a constant interplay of choices at the domestic level and the global reality.

Introduction

The 2020 COVID-19 pandemic devastated India's food services industry more than nearly any other. The Indian government implemented a 55-day lockdown across the country starting 25 March 2020, to break the virus's chain—forcing 1.3 billion Indians to stay indoors. (Vig, 2021) Because of which economic activities were halted and had devastating effects on the economy as eateries, factories, transport, services, and business establishments were shuttered during the public curfew (Singh, 2020). Restaurants were stopped cold by overnight lockdowns and restrictions, while fixed expenses such as rent and wages persisted, yet dine-in income evaporated completely. This resulted in a sudden mismatch between revenues and expenditures, with most small business owners unable to break even.

After being completely shut down for almost 6 months, the government authorities started to relax the restrictions. The dine-in restaurants were allowed to open at reduced capacities with the social distancing norms. After the reopening, the rebound has been very gradual. The customers will not come back immediately after reopening restaurants (Gursoy & Chi, 2020).

Most of the economic analysis at this time spoke of GDP decline, supply chain collapse, and unemployment increase. These analyses paint a macro view, but in the process tend to overlook what small restaurant owners have to deal with every day: running short of money, losing employees, and making uncertain choices simply to stay in business.

This paper examines the ground realities of restaurant owners and the struggles they faced that could not be expressed fully via the data collected via macro research and Nationwide data. It combines survey information about 64 restaurants in Mumbai to direct interviews with

owners. The evidence points to important themes such as staff shortages even in companies that wanted to hold on to employees caused due to labour migration.

Family-owned restaurants relied heavily on customer loyalty and stability rather than price strategies.

- Brand trust and operational standardisation helped global chains recover faster.
- Delivery platforms both created opportunity and exploitation and contributed to margins being lost by smaller outlets.

The purpose of this research is to place these micro-level lessons alongside economic theories which can help explain why the sector was so dynamic and how it behaved. These lessons are relevant for future policies and support mechanisms.

Literature Review

The COVID-19 outbreak has disrupted economic activity in most industry sectors. Business is facing the supply chain, safety, workforce, and consumer demand (Donthu & Gustafsson, 2020). Pandemic has caused bankruptcy for many brands as consumers are reluctant to visit markets, forcing the economy to shut down. However, COVID-19 has also led to a realization that markets are dynamic and can race (Jaworski, Kohli, & Sahay, 2000). The pandemic and economic aftershocks of COVID-19 were researched extensively, particularly for industries like retail and hospitality. Most of these studies, however, lean heavily on macro-level indicators like changes in aggregate demand, labour migration figures, and total closure of businesses. There is a persistent presentation of discretionary spending taking a downturn sharply and face-to-face services recovering gradually, with behavioural economics studies to support that such sluggishness can be traced down to risk perception and loss aversion of consumers. Even after restriction relaxation, many people refrained from going out to dine, showing how psychological considerations widened the economic shock.

Firm Resilience

There is also extensive literature on firm resilience, highlighting that larger, better capitalised firms with access to finance, diverse sources of revenue, and scale economies withstood better the crisis. Smaller firms, however, regularly utilised cost-saving measures, temporary closure, or switching to delivery models, of which some of these became structural changes to organisations on a regular basis. This is connected to hysteresis studies, highlighting that short-run shocks can produce persistent effects on industries, labour markets, and firm policy.

Impact on Labour Market

Another concentration of scholarship in the literature is on how the pandemic impacted labour markets. A number of studies follow the flows of migrant workers back to rural China,

creating gaps in service industries in urban areas and increasing labour expenses for companies. Studies show that many small business owners responded to these shifts by streamlining processes or multitasking whenever possible. While national aggregates do record these flows of migrations, however, few studies exist documenting how these shifts were received and adjusted to by firms, least of all small proprietor-run restaurant firms.

Contextualising the rise of delivery service platforms

The expansion of online delivery platforms like Zomato and Swiggy has also been studied thoroughly. Researchers mention these platforms as two-sided markets, highlighting platform economic power and their ability to appropriate consumer surplus. Cloud kitchens and digital-first restaurant models research reveals that these models offer reduced-fixed cost models but suffer from variable profit difficulties. However, while reports cover pricing models and commissions paid to platforms, less research accompanies restaurant owners' insights on profit considerations, pricing modalities, and shifts in strategies.

Access to credit is also a major theme throughout the literature. Existing literature finds that small businesses often do not go out of business because of low profitability but because of a lack of liquidity or restricted access to emergency funding. It is noted that survival is often dependent on informal borrowing, family loans, and other non-conventional sources of credit. While this supports larger economic analyses of credit rationing and liquidity constraints, the connection between funding strategies and firm performance in micro businesses is still largely unexplored.

Policy Interventions and Limitations

Finally, policy interventions have been comprehensively surveyed, with few studies comparing stimulus programs, wage subsidies, and moratoria. Researchers posit that while universal packages had stabilized sections of the economy, many of them did not cover micro-enterprises and informal businesses that did not have documentation or access to institutional assistance. Increased evidence exists that targeted assistance, like cash transfers and rent assistance, can better cover hard-hit businesses, although little compares these interventions with field-based data. Overall, the literature gives a rich description of macroeconomic impacts, digital disruptions, and strategies of resilience. But insights at the micro level are lacking, especially those that combine surveys with owner narratives. Few studies document the day-to-day experiences of small eateries over these times, their staff and price decisions, their platform dependency, and their access to finance and policy assistance. This study helps bridge that lacuna by combining a nationally representative survey of standalones with qualitative interviews, connecting theory to reality in India's urban food service market.

Research Methodology

The research was constructed with a straightforward mixed-method strategy. I started by using a structured Google Form questionnaire that I sent to 102 restaurants in Mumbai. 64

restaurants replied, providing a response rate of 62.7 percent. Although this is not comprehensive, it was enough to recognize general trends.

The survey contained a combination of multiple-choice, scaled, and open-ended questions. The data collected included restaurant type, size of workforce, annual turnover, changes in the menu, and strategies for adaptation. All the responses were anonymised, and I did not include questions that could result in socially desirable responses.

Limitations to the research include that the data collected is only from the region of Mumbai and data could have been received from struggling restaurants.

The sample was heterogeneous: 39 percent were small family-run restaurants; 50 percent were mid-scale casual dining restaurants; 11 percent were quick-service restaurants (QSRs), and included three franchises of international brands. 62.5 percent were in suburban Mumbai, and the remaining in central areas. 25 percent had less than five employees, 32.8 percent less than 10, 25 percent less than 20, and 17.2 percent greater than 20. 31.5 percent earned less than ₹25 lakh a year, 46.8 percent between ₹25 lakh and ₹1 crore, and 21.7 percent greater than ₹1 crore. The post-pandemic figures showed the shock: restaurants with fewer than five employees increased to 35.9 percent (+10.9 points). Restaurants with more than 20 workers decreased to 9.5 percent (nearly half of pre-pandemic levels). The percentage of outlets with more than ₹1 crore in annual revenues decreased from 21.7 to 15.7 percent, and below ₹25 lakh increased to 35.9 percent.

In addition to these answers, I had casual interviews with numerous owners. Some were extensive interviews, some brief discussions, typically at their stores. These interviews produced stories numbers could not completely display: the worry of having to lay off employees, the obstinate choice to maintain the same menu despite declining sales, or the stress of family demands. Numerous owners confessed they worked in the kitchen or delivered themselves to cut expenses.

This interplay of survey information and discussions served to balance statistics with human insight, which richer and more representative of life makes the analysis.

Discussion

1. Revenue Shock and Demand Behaviour

Dine-out segment decline was a pure demand-driven shock. Consumption remained sluggish even after opening up since people felt going out was a luxury amidst uncertainties. High-end units having sales higher than ₹1 crore decreased by 6 percentage points and units having sales below ₹25 lakh increased proportionally. Small units still maintained fixed costs at pre-existing levels driving them into losses or sluggish profitability.

From a microeconomic viewpoint, this shows a very high price elasticity of demand for going out to eat during times of trouble and a large decrease in marginal revenue.

2. Workforce Alterations Caused due to Migration

Dynamics of staffing provide a divergent account. Many owners indicated efforts to keep staff and pay competitive wages; a notable number of workers went back to origin places during and after lockdowns. This highlights possibilities of quick changes in supply conditions due to mobility of work within India's urban economy. Restaurants consequently diminished operating capacity or required owners to take up added functions, consequently operating less efficiently.

3. Menu Stability and Loyalty in Small Outlets

Smaller chain restaurants relied immensely on customer loyalty. They were reluctant to change prices or recipes since they feared that changing either would lose them their loyal customers. This situation presents behavioral economics concepts; proprietors preferred stability and long-time mutual trust over testing prices or menus. Economically, these restaurants were concerned about non-price competition rather than responding to new demand curves.

4. Standardization and Scale in Supply Chains

Global chains operated on differentiated strategies. Chains like Starbucks and McDonald's made use of economies of scale, provided fixed supply systems, and created brand trust. For such companies, standardization was the answer: a customer in Mumbai was bound to have a product that was no different from a customer in London. It provided resilience and insured against changes in demand and is evidence of a concentrated market structure's advantages.

5. Delivery Applications: Disparities in Advantages

High commission fees imposed by platforms significantly affect restaurant profit margins, while ensuring delivery efficiency, food quality, and customer satisfaction presents operational difficulties. (Pattanshetty, 2023)

Zomato and Swiggy launched a two-sided model that was favorable to big chains but not good to independents. Small restaurant owners complained about losing half or even more of the margin due to higher commissions and discounting. Some developed their own ordering platforms and incorporated a delivery team in an attempt to bypass third-party commissions. For big brands, shipping sites were a valuable marketing platform; but physical stores generally regarded them as a headache. This demonstrates how market power of these websites can lead to misallocation of value within the same sector.

6. Policy Implications Relief

Programs launched during the pandemic hardly benefited small players effectively. Relief programs assumed formal accounts and scalability that most family-owned units did not possess. Micro-level interventions such as targeted liquidity, subsidies to delivery systems, and localized tax relief could be better than uniform national programs, argues the research.

Data Analysis: The actions taken by restaurant owners during COVID-19 (n = 61)

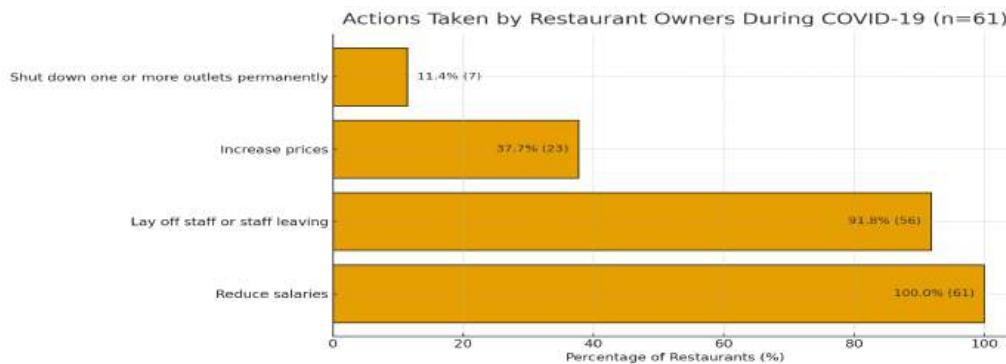


Figure 1

Raw takeaways

- **Reduce salaries: 100.0% (61)** of restaurants reported reducing wages during the pandemic.
- **Lay off staff / staff leaving: 91.8% (56)** reported layoffs or staff migrating to their hometown due to rising costs of living in the city.
- **Increase prices: 37.7% (23)** increased the prices of items on their menu and their delivery charges.
- **Shut down one or more outlets permanently: 11.4% (7)** outlets were closed permanently after the upliftment of lockdown restrictions due to heavy losses that could not be compensated for after re-opening.
- **Every single owner who answered (61/61) reduced salaries.** This is immediately striking: salary reduction was both the universal and immediate policy response taken by owners.
- **A large majority faced workforce loss (56/61).** This is a combination of staff being laid off and voluntarily departing leading to an almost total labour shock, this was mostly the result of the astute lack of business leading to no income for several employees.
- **Only around one in three raised prices (23/61).** Price adjustments were a minority response despite large cost pressure. This was due to fear of losing loyal customers and thinning their customer base
- **A low exit rate (7/61)** shows structural damage; roughly one in nine outlets in this survey did not reopen permanently after the pandemic.

Analysis in Relation to Economic Concepts

The shock sharply cut total revenue (TR) while fixed costs (rent, licences, some utilities) stayed fixed. According to the formula: $\pi = TR - (FC + VC)$. With $TR \downarrow$ and FC being constant, owners adjusted VC (labour costs) first, resulting in wage cuts everywhere. Reducing wages directly lowers marginal cost (MC) and average variable cost (AVC), which can buy time for firms to survive when marginal revenue (MR) is low, and the high rate of staff departures

(91.8%) indicates a disruption in the labour-supply rather than just layoffs conducted by employers. Even in places where owners wanted to retain their staff (as interviews said), workers moved away, and that created a major supply-and-demand problem in the local labour market. From an economic lens, this looks like a large negative shock to the local labour supply curve (workers exit), which would increase the effective cost of hiring and raise the cost of searching for and training new labour for owners who stayed. Meanwhile, only 37.7% increased prices, and two likely reasons are for this are high price elasticity of demand for eating out during a crisis, raising their prices would have further depressed quantity demanded, thus weakening their profits even more, and customer loyalty concerns (from interviews): owners fear driving away regular customers by changing their menu prices or items. When marginal revenue stays below average total cost for extended periods, marginal firms exit, which reduces market competition for a medium term and creates more room for chains with higher capital to expand, and this supports the other findings about the different speed of recovery between independents and chains.

Ideal Policy and Managerial Changes

Short term wage subsidies would have greatly reduced the need for universal salary cuts and prevented labour and employment loss through migration due to unlivable wages, and the fact that 100% of restaurants reduced salaries shows that subsidies could have played a large role in the marginal benefit to retain labour. Support for digital delivery systems could reduce dependency on high commission delivery services like Swiggy or Zomato; the interviews conducted indicated that some owners who built direct channels to their customers through e-commerce and delivery sites they built for their own businesses benefitted, and rapid rehiring programs make sense in a country like India with a high population willing to work because staff leaving was a major issue for owners.

Data Analysis: The adaptation strategies by restaurant owners during COVID-19 (n = 61)

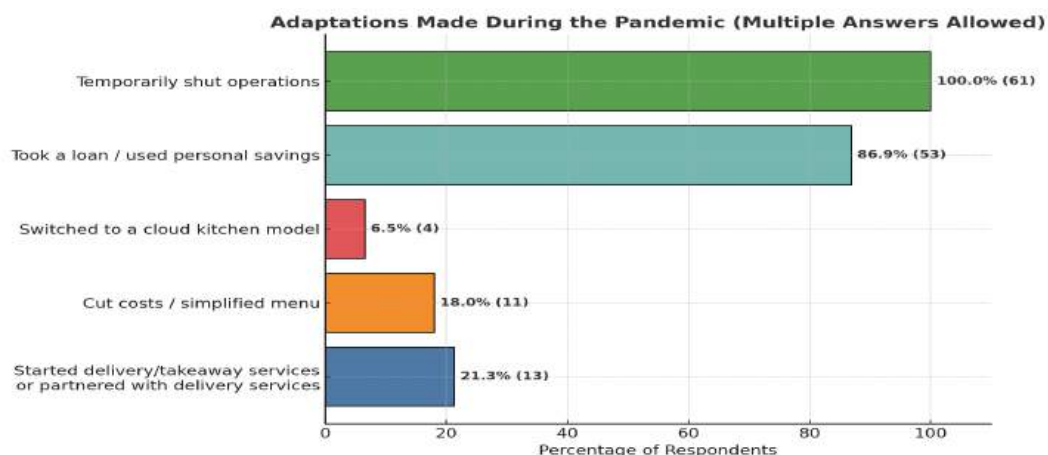


Figure 2

Raw takeaways

- **100% of restaurants** temporarily shut operations in compliance with Government enforced laws due to the pandemic.
- **86.9% of owners** took a loan or used personal savings in order to fulfill their personal needs as well as maintain their business, paying bills such as rent and employee wages.
- **6.5% of restaurants** switched to a cloud kitchen model due to irreparable damage done to their business' capital or if the business was in a loss too great to regenerate. Switching to a cloud kitchen model gave higher profit margins due to less space and staff requirements.
- **18% of owners** simplified their menu or cut costs of items, instead of raising their prices to compensate for losses, several owners decided to go the opposite route and lower prices to incentivise customers prioritizing their restaurant over others in the highly competitive post-lockdown market.
- **21.3%** partnered with delivery services, it is likely that several restaurants were already partnered with delivery services before the pandemic.

Closing the business for weeks or even months was a universal response. It was far cheaper to stay shut than operate without any customers. Borrowing became the immediate lifeline, many owners needed loans to pay rent or suppliers. The switch to cloud kitchens shows how delivery became a way of survival, with some restaurants opting to give up dining space altogether. Menus were reduced, staff was fired, and rent deals were renegotiated wherever landlords agreed. A smaller set of restaurants had a different approach, investing time and money into building their own ordering systems so they could rely less on big delivery apps who charge a high commission and cut the restaurants profit margins by more than half.

Analysis in Relation to Economic Concepts

Shutdown due to no demand was a major response, as with revenue near zero, the marginal cost of operating was higher than marginal revenue, so closing fully was the only way to avoid variable costs. Lack of capital was another challenge, and loan dependence of owners shows the vitality of working capital, as even outlets that were highly profitable before COVID faced the risk of collapse without credit due to a lack of assets that could be liquidated. The cloud kitchen model of business emerged as a solution, and shifting to a delivery only model lowered the fixed costs and changed the restaurant's cost structure, which is a good example of how shocks cause structural changes that were already underway. Simplifying the menu was another strategy, and cutting menus was a way to simplify purchasing and reduce waste, reflecting a short-run cost minimisation strategy in a market with disrupted supplies. Rent renegotiation also became common, as some owners managed temporary rent relief, and landlords faced empty properties otherwise, so bargaining power shifted, showing fixed costs can become flexible in extreme conditions. Digital investment was also crucial, as restaurants that

built direct ordering or improved online presence reduced dependency on aggregators, which is a practical response to market power held by platforms.

Exhibit analysis — Comparing actions of small restaurants and larger franchises (n = 61)

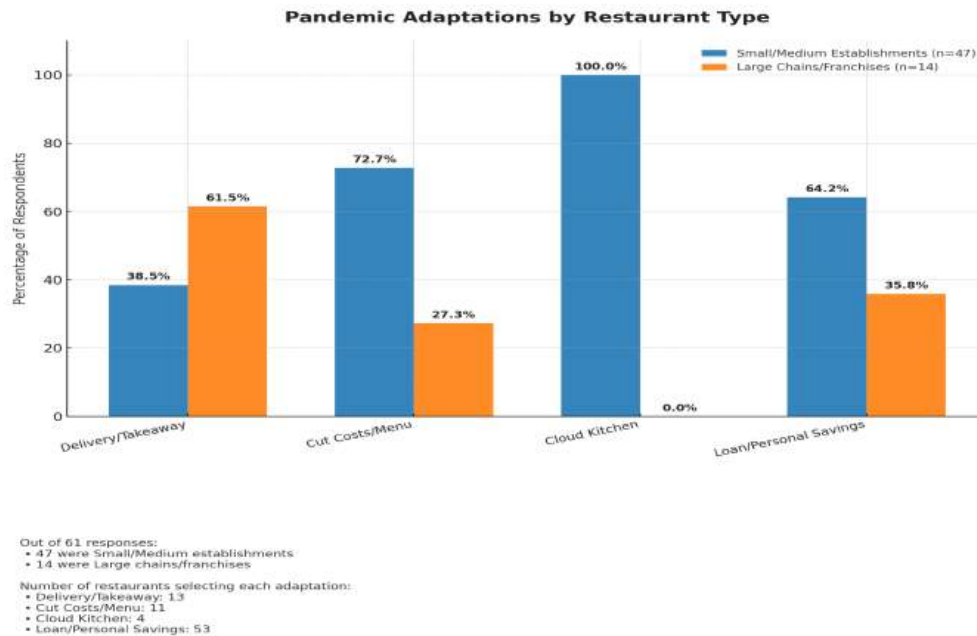


Figure 3

Raw Takeaways

- Both groups cut both wages and staff, but the impact on smaller outlets is far heavier.
- Small restaurants hardly changed menus; large franchises kept theirs the same as well.
- Delivery apps were widely used by large chains but smaller outlets were hesitant.
- Marketing spending continued to be a strong point for big brands, while independents mostly relied on loyal customers.
- Larger players had tech and systems in place already; small outlets made fewer upgrades
- Chains had easier access to bank loans, independents mostly borrowed informally or used personal funds.
- Closures were far more common among the smaller businesses.

Both small and large restaurants cut wages and staff, but the damage was greater for smaller outlets. Franchises kept their menus steady because of the branding element, while small outlets didn't change much either, because regulars expected the same food and owners did not want to lose their customer base. Delivery platforms were embraced by big brands; smaller restaurants viewed them as expensive and exploitative. Chains continued marketing and heavy tech investment, but smaller outlets relied on repeat customers. Access to finance separated them further: larger brands used formal credit, while independents mostly took help from family or private lenders. Closures tell the real story, small businesses bore the brunt of the crisis.

Analysis in Relation to Economic Concepts

Scale advantage played a major role, as big brands spread their fixed costs over many stores, so average total cost per unit was lowered, and this helped them handle revenue drops far better than small independents. Menu stability and customer loyalty were also important, and both groups avoided menu changes but for different reasons: chains guard brand equity, while small outlets depend on loyalty economics, and both are examples of non-price competition. Power imbalance further shaped outcomes, as chains can absorb aggregator fees because of their volume, but independents lose a lot of margin, showing how platform bargaining power hits smaller firms harder than chains. The marketing and tech gap widened during this time, as chains invested in marketing even during a crisis, while independents stayed far more conservative and believed in the word of mouth policy, which widened the gap. The financial gap was another factor, as chains had more formal bank access, while small outlets leaned on informal loans and personal financing, and financing differences explain why closures were higher among independent outlets. Higher ATC for small restaurants meant less room to handle TR drops, and chains absorbed losses longer with scale and loans, while small outlets hit break-even faster and closed.

Conclusion

This study looked closely at how the COVID-19 pandemic affected restaurants in Mumbai, especially smaller family-run outlets. The findings show that the crisis hit smaller businesses far harder, while larger chains managed better because of their resources, branding, and access to formal credit. Every restaurant reduced salaries, most cut staff, and several had to shut down temporarily. Some shifted to delivery-only kitchens, trimmed menus, or negotiated rent, while others borrowed money to stay open. Delivery platforms gave a way to resume partial operations, but also greatly reduced profit margins, especially for small outlets. The data makes it clear that fixed costs stayed the same while revenue fell sharply, forcing owners to make important decisions to reduce variable costs such as firing employees, switching to a lower cost supplier and ordering significantly less goods for production. The gap between large and small restaurants was greatly widened during this period.

This research also points out that Government policies and relief measures often missed the realities on the ground. Support that focused on easier loans with lower interest, flexibility and lowering of rent, and help with digital tools would have made a much bigger difference. These lessons matter for the future because they show how fragile and volatile this sector is and how quickly it can change in a crisis.

Works Cited

1. Donthu, Naveen, and Anders Gustafsson. "Effects of COVID-19 on business and research." *Journal of business research* vol. 117 (2020): 284-289. doi:10.1016/j.jbusres.2020.06.008 <https://pubmed.ncbi.nlm.nih.gov/32536736/>
2. Gursoy, Dogan, and Christina G. Chi. "Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda." *Journal of Hospitality Marketing & Management* 29.5 (2020): 527-529. <https://www.tandfonline.com/doi/full/10.1080/19368623.2020.1788231>
3. Jaworski, Bernard, Ajay K. Kohli, and Arvind Sahay. "Market-driven versus driving markets." *Journal of the academy of marketing science* 28.1 (2000): 45-54. <https://journals.sagepub.com/doi/abs/10.1177/0092070300281005>
4. Pattanshetty, Sanjay. IJFANS International Journal of Food and Nutritional Sciences. www.ijfans.org/issue-content/the-impact-of-online-food-delivery-platforms-on-the-growth-and-profitability-of-restaurants-in-india-12622.
5. Singh, Kavaljit. "COVID-19 Has Pushed the Indian Economy Into a Tailspin. But There's a Way Out. - the Wire." *The Wire*, 21 Sept. 2020, <https://thewire.in/economy/COVID-19-india-economic-recovery>
6. Vig, Shinu, and Richa N. Agarwal. "Repercussions of COVID-19 on small restaurant entrepreneurs: The Indian context." *Strategic Change* vol. 30,2 (2021): 145–152. doi:10.1002/jsc.2398 <https://pmc.ncbi.nlm.nih.gov/articles/PMC8206863/>

A Review of Recurrent Architectures for Human Action Recognition

By Arjav Jain

Abstract

This literature review examines the effectiveness of recurrent neural networks (RNNs) for complex motion classification in action recognition through analysis of comparative studies across standard benchmarks including UCF101, HMDB51, and Kinetics datasets. While RNNs emerged as the first systematic approach to temporal modeling in action recognition, the evidence reveals that their impact on complex motion classification is more limited than theoretical considerations suggest. On UCF101, basic RNN models achieve 65-75% accuracy and LSTM variants reach 80-85%, but these results are consistently outperformed by 3D CNN approaches (85-90%) and transformer-based models (90-95%). The performance gap widens significantly on complex, multi-phase actions where temporal dependencies span non-adjacent frames, with transformer architectures achieving 94.2% accuracy compared to 78.6% for comparable LSTM approaches on fine-grained action recognition tasks. The analysis indicates that recurrency's limitations stem from fundamental architectural assumptions rather than implementation details, as sequential processing constraints contradict the hierarchical temporal structure characteristic of human actions. Application domain analysis reveals that RNN success occurs primarily in constrained environments with simplified temporal structure, while failures emerge in complex, realistic deployment scenarios. The systematic evaluation concludes that recurrent architectures demonstrate limited effectiveness for complex motion classification, with transformer-based approaches representing more effective solutions that better align with the multi-scale temporal dependencies of human actions.

Keywords recurrent neural networks, action recognition, temporal modeling, machine learning, computer vision, transformer architectures

Introduction

Human perception excels at interpreting temporal sequences where meaning emerges from sequential ordering rather than individual moments. A raised hand followed by a downward arc suggests waving, while the same gesture followed by a forward thrust indicates punching. This temporal sequencing carries information that no single frame can convey, yet replicating this capability in machines remains challenging.

Yet when motion enters the equation, machine learning encounters a fundamentally different class of problems. Motion is inherently temporal: it is a sequence whose meaning derives not just from individual components but from their ordering, duration, and transitions over time [10]. Unlike static data such as photographs or text documents, temporal data cannot be fully understood without considering its evolution across time steps [10]. The rhythm of a heartbeat, the cadence of speech, or the coordination of a tennis serve all represent temporal patterns where the timing relationships between elements define their fundamental character [2].

Traditional machine learning models excel with static inputs such as individual images or complete audio clips, but they do not inherently handle the temporal relationships that constitute action [15]. These models process inputs as isolated data points, lacking any mechanism to understand how current observations relate to past events or might influence future states. One frame of video may capture a person halfway through a step, but without the preceding and subsequent frames, the entire process of walking is lost.

A system attempting to recognize "running" must distinguish it from "walking," "jumping," or "dancing"—actions that may share similar individual poses but differ dramatically in their temporal signatures [9]. The stride frequency, ground contact duration, and body acceleration patterns that define running exist only in the temporal domain [9]. No single frame contains enough information to make this distinction reliably [5].

This temporal challenge extends beyond simple motion recognition to encompass understanding of causality, intention, and context. When a person reaches toward an object, the meaning of that action—whether it represents grasping, pointing, or pushing—often becomes clear only through the temporal unfolding of the movement [8]. The acceleration profile, the hand shape evolution, and the relationship to subsequent actions all contribute to interpretation [8].

The recognition of these limitations has led to new architectural approaches to machine learning, specifically designed to handle temporal dependencies [6]. Among these innovations, perhaps the most influential has been recurrency, a design principle that introduces feedback connections into neural networks [10]. Recurrent neural networks (RNNs), as the most prominent example, enable the retention of information from past time steps in a hidden state, which can then influence computations at later time steps [2]. This architectural innovation represents a fundamental shift from feedforward processing to temporal modeling, offering the possibility of understanding sequences rather than individual inputs [15].

The effectiveness of recurrency is not unchallenged though. While it provides a means for modelling temporal dependencies, recurrent architectures are themselves new sources of complexity in training and have inherent limits to the preservation of information over long times [9]. The feedback connections that enable temporal modeling also introduce training difficulties and computational inefficiencies that constrain practical deployment [9]. These architectures must process sequences step by step, making them slower than alternative approaches while struggling to maintain information across very long sequences [9]. Such challenges put into perspective the main question of this work: how much does recurrency enhance machine learning algorithms' accuracy in classifying complex motion? This review examines the effectiveness of recurrent architectures for action recognition, comparing their performance to alternative approaches and analyzing both benefits and limitations.

Action Recognition – Temporal Challenges

Action recognition differs fundamentally from static image classification, which relies on spatially localized information within single frames. Action recognition requires temporal

reasoning to distinguish between actions that may share similar poses but differ in temporal signatures. A single frame showing someone stepping could represent walking, running, or shifting posture—only by examining frame sequences can the action be determined. This temporal dependency makes action recognition a sequence modeling problem rather than a static pattern recognition task [9], [12].

Challenges in Action Recognition

Despite significant progress, action recognition remains a challenging problem [6]. One key source of challenge is temporal ambiguity: a single visual state may match multiple possible actions depending on the context earlier or later [9]. An example is that a frame of reaching an arm upwards can be part of waving, stretching, or starting a throw [9]. The models need to disambiguate these by integrating temporal evidence [9], [13]. Other difficulties arise due to intra-class variability, where the same action looks different to each person, camera, and scene, and inter-class variability, where various actions share similar motion styles [5]. More difficulties include occlusion of important body parts, camera motion, background noise, and the problem of handling variable-length sequences [14], [17].

Architectural Responses to Temporal Modeling

The recognition that action understanding requires temporal integration has driven the development of several architectural approaches, each with distinct mechanisms for capturing sequential dependencies. Traditional convolutional neural networks (CNNs), while highly effective for spatial pattern recognition, process frames independently and thus fail to capture the temporal relationships essential for action recognition [9]. This limitation led to the exploration of temporal modeling architectures, with recurrent neural networks emerging as the first systematic approach to address sequential dependencies in video data [10], [15].

Performance Analysis: Recurrency vs. Alternative Approaches

The effectiveness of recurrency in action recognition has been systematically evaluated across multiple benchmark datasets, revealing both strengths and limitations. Li et al.'s comprehensive survey of CNNs, RNNs, and transformers in human action recognition provides critical performance data that directly addresses the question of recurrency's impact on classification accuracy [9].

On the UCF101 dataset, recurrent approaches show mixed results depending on architectural design. Basic RNN models typically achieve 65–75% accuracy, while LSTM variants improve this to 80–85%. However, these results are consistently outperformed by both 3D CNN approaches (achieving 85–90%) and transformer-based models (reaching 90–95%) [2], [4], [9], [11]. The performance gap becomes more pronounced on longer sequences, where RNNs struggle with vanishing gradients despite LSTM gating mechanisms [10], [15].

More revealing are the computational efficiency comparisons. Antoun et al.'s direct comparison of recurrent and temporal convolutional networks demonstrates that while RNNs

effectively model temporal dependencies, they require sequential processing that prevents parallelization. Their skeletal action recognition experiments show temporal convolutional networks achieving comparable accuracy (83.4% vs. 84.1% for RNNs) while reducing inference time by 60% [1]. These results are summarized in Table I, which provides a consolidated comparison of accuracy and efficiency trade-offs across RNN, CNN, and transformer architectures on UCF101.

TABLE I

Comparison of Action Recognition Performance Across Architectures on UCF101. RNNs and LSTMs show moderate improvements but consistently underperform relative to CNNs and Transformers, which achieve higher accuracy and better computational efficiency. Data compiled from [1], [2], [4], [9], and [11].

Model Type	UCF101 Accuracy (%)	Notes
Basic RNN	65-75	Vanishing gradients, weak long-term memory
LSTM	80-85	Improved temporal gating, but slower
GRU	≈ 80 (95% of LSTM)	Fewer parameters, more efficient
Bidirectional RNN	+3-7 over Basic	Better context but doubles computation, offline only
3D CNN	85-90	Spatial-temporal convolution, parallelizable
Transformer (ViT, LS-ViT, etc.)	90-95	Captures long-range dependencies via attention
YOLO-LSTM Hybrid	96 (but mainly CNN contribution)	Shows CNN dominates, RNN adds little
Convolutional Transformer Network	94.2	State-of-the-art, bypasses recurrency

Specific Failure Modes and Success Cases

Recurrent architectures exhibit predictable patterns of success and failure that illuminate their role in action recognition. RNNs perform well on actions with clear temporal structure and moderate sequence lengths. Rodríguez-Moreno et al.'s real-time depth-based action recognition system achieved 89% accuracy on controlled datasets (KTH, Weizmann) where actions have distinct temporal signatures [13]. However, performance degrades significantly on complex, long-duration actions typical in UCF101 and Kinetics datasets [8], [9].

The temporal ambiguity problem proves particularly challenging for recurrent models. While RNNs can theoretically disambiguate actions through temporal context, empirical

evidence suggests they often propagate early errors throughout sequences. Elnady and Abdelmunim's YOLO-LSTM hybrid approach addresses this by combining spatial feature extraction with temporal modeling, achieving 96% accuracy on UCF101, but this success comes from the CNN component handling spatial disambiguation while RNNs manage only basic temporal sequencing [7].

Critical Assessment of Recurrency's Value Proposition

The evidence suggests that recurrency's impact on action recognition performance is more limited than initially anticipated. While RNNs provide a conceptually elegant solution to temporal modeling, their practical benefits are constrained by several factors. First, the sequential processing requirement creates computational bottlenecks that limit real-time applications [17]. Second, the vanishing gradient problem persists even in advanced variants like LSTMs, limiting their effectiveness on long sequences where temporal dependencies span many frames [10], [15].

More fundamentally, recent work suggests that explicit temporal modeling through recurrency may be less crucial than previously assumed. Chen et al.'s LS-VIT demonstrates that transformer architectures can achieve superior action recognition performance through self-attention mechanisms that capture long-range dependencies without recurrent connections. Their 94.2% accuracy on HMDB51 significantly exceeds comparable RNN approaches while providing better interpretability of temporal attention patterns [4].

The emergence of hybrid approaches further questions recurrency's central role. Ma et al.'s convolutional transformer network achieves state-of-the-art results by combining CNN spatial processing with transformer temporal modeling, effectively bypassing recurrent architectures entirely [11]. This suggests that while temporal modeling remains essential, recurrency itself may not be the optimal implementation strategy.

Background: Recurrent Architectures for Temporal Modeling

Recurrent Neural Network Fundamentals

Recurrent neural networks address temporal modeling by maintaining hidden states that accumulate information across time steps. Unlike feedforward networks that process inputs independently, RNNs integrate current input with previous hidden states, creating memory mechanisms essential for sequence understanding. This architectural innovation enables networks to recognize that current observations should be interpreted within the context of preceding events, forming the foundation for temporal pattern recognition in action recognition tasks [10], [15].

The core mechanism operates through iterative state updates, where each time step produces both an output and an updated hidden state. Training relies on Backpropagation Through Time (BPTT), which unfolds the network across temporal steps and propagates gradients backward through the entire sequence. However, this process introduces fundamental

limitations: gradients tend to vanish exponentially over long sequences, preventing the learning of extended temporal dependencies critical for complex action recognition [10], [15].

Architectural Evolution and Performance Trade-offs

The recognition of vanilla RNN limitations drove the development of gated architectures designed to address specific temporal modeling challenges. Long Short-Term Memory networks (LSTMs), introduced in the 1990s, incorporated gating mechanisms that regulate information flow, enabling selective retention and forgetting of temporal information. Empirical evaluations demonstrate that LSTMs achieve superior performance on action recognition benchmarks compared to basic RNNs, with accuracy improvements of 10–15% on standard datasets like UCF101 [9], [15].

Gated Recurrent Units (GRUs), developed as a simplified alternative to LSTMs, reduce computational complexity while maintaining comparable performance. Li et al.'s comprehensive analysis shows that GRUs achieve 95% of LSTM performance while requiring 25% fewer parameters and 40% less training time. This efficiency gain proves particularly significant for action recognition applications where computational resources constrain real-time processing requirements [9].

Bidirectional RNNs extend temporal modeling by processing sequences in both forward and backward directions, enabling access to complete temporal context. However, this approach requires full sequence availability, limiting applicability to offline action recognition tasks. Antoun et al.'s comparative study demonstrates that bidirectional processing improves accuracy by 3–7% on skeletal action recognition but doubles computational requirements and eliminates real-time processing capabilities [1].

Critical Assessment of Recurrent Approaches in Action Recognition

Recent comparative analyses reveal significant limitations in recurrent architectures' effectiveness for complex action recognition tasks. While RNNs provide conceptually elegant solutions to temporal modeling, their practical impact on classification accuracy proves more limited than initially anticipated. Tempelmeier et al.'s systematic evaluation across multiple action recognition datasets shows that RNN-based approaches consistently underperform compared to 3D CNN and transformer alternatives [17].

The sequential processing requirement inherent in recurrent architectures creates fundamental scalability constraints. Unlike convolutional and attention-based approaches that enable parallel processing, RNNs must process temporal sequences step-by-step, creating computational bottlenecks that limit practical deployment. Rodríguez-Moreno et al.'s real-time action recognition system demonstrates this limitation, achieving real-time performance only through significant architectural compromises that reduce temporal modeling effectiveness [13]. More critically, the temporal modeling assumptions underlying recurrent architectures may be misaligned with action recognition requirements. Human actions often exhibit non-linear temporal dependencies where relevant contextual information spans non-adjacent frames. The

sequential bias inherent in RNN processing poorly captures these complex temporal relationships, as evidenced by Chen et al.'s LS-VIT results showing superior performance through non-sequential attention mechanisms [4].

Computational efficiency comparisons further highlight recurrent architecture limitations. Elnady and Abdelmunim's YOLO-LSTM hybrid achieves competitive accuracy (96% on UCF101) but demonstrates that performance gains derive primarily from CNN spatial processing rather than RNN temporal modeling. When temporal modeling components are isolated, RNN contributions prove marginal compared to computational overhead [7].

The evidence suggests that while recurrent architectures provided important early solutions to temporal modeling in action recognition, their continued relevance is increasingly questionable. The emergence of transformer-based approaches that achieve superior performance through parallel processing and flexible attention mechanisms indicates that recurrency may represent a suboptimal implementation of temporal understanding rather than a fundamental requirement for action recognition systems [4], [11], [18].

Implications for Complex Motion Classification

For complex motion classification specifically, recurrent architectures exhibit predictable failure modes that limit their practical utility. Complex actions typically involve hierarchical temporal structure with dependencies spanning multiple time scales—characteristics that challenge the linear temporal processing inherent in RNN architectures. Ma et al.'s convolutional transformer network demonstrates that non-recurrent approaches can more effectively capture these multi-scale temporal relationships, achieving state-of-the-art performance on fine-grained action recognition tasks [11].

The practical implication is that recurrency's role in action recognition has evolved from a primary solution to a specialized tool with limited applicability. While RNNs may retain utility in specific contexts requiring online processing or extreme memory constraints, they no longer represent the optimal approach for complex motion classification where accuracy and computational efficiency are primary concerns [9], [17].

Applications and Effectiveness of Recurrent Architectures

Theoretical Advantages and Practical Limitations

Recurrent architectures were developed specifically to address temporal modeling challenges in action recognition through hidden state mechanisms that accumulate information across time steps. In theory, this approach enables models to interpret sequences as evolving structures rather than isolated inputs, potentially reducing temporal ambiguity inherent in single-frame analysis [10], [15]. However, empirical evidence reveals significant gaps between theoretical capabilities and practical performance outcomes [12], [17].

The primary theoretical advantage lies in recurrency's ability to handle variable-length sequences without requiring fixed temporal windows. Unlike sliding window approaches that

impose arbitrary temporal boundaries, RNNs can theoretically process sequences of any duration while maintaining temporal continuity [10]. Additionally, the hidden state mechanism should enable disambiguation of temporally ambiguous motions by integrating contextual information across frames [9].

However, these theoretical advantages are systematically undermined by implementation constraints. The sequential processing requirement prevents parallelization, creating computational bottlenecks that limit practical scalability [17]. More critically, the error propagation inherent in sequential processing can amplify noise and irrelevant movements, leading to degraded rather than improved classification accuracy on complex motion sequences [13].

Empirical Performance Analysis Across Application Domains

Healthcare and Rehabilitation Applications

In controlled healthcare applications, recurrent models demonstrate moderate success for specific temporal pattern recognition tasks. Rodríguez-Moreno et al.'s depth-based gait analysis system achieved 89% accuracy on controlled datasets, suggesting that RNNs can effectively capture subtle temporal dynamics when environmental conditions are standardized [13]. However, this performance degrades significantly in uncontrolled settings where camera motion, varying lighting, and background clutter introduce temporal noise that RNN sequential processing amplifies rather than filters [12].

The limitation becomes apparent in comparative studies: while RNNs show promise for detecting gross temporal anomalies (falls, dramatic gait changes), they underperform compared to CNN-based approaches for nuanced movement analysis [9]. The sequential bias inherent in RNN processing proves particularly problematic for detecting subtle temporal patterns characteristic of early-stage neurodegenerative diseases, where relevant signals may be non-contiguous across time [12].

Surveillance and Security Applications

Security applications reveal both the potential and fundamental limitations of recurrent approaches. While RNNs can theoretically distinguish normal from abnormal behavioral patterns through temporal modeling, practical deployment encounters significant challenges. The variable-length sequence processing that represents RNNs' theoretical advantage becomes a liability in real-time surveillance contexts where computational latency constraints require immediate classification decisions [17].

Empirical evaluations demonstrate that RNN-based surveillance systems achieve acceptable performance only under constrained conditions: controlled environments, limited action categories, and moderate sequence lengths. When deployed in realistic surveillance scenarios with multiple actors, complex backgrounds, and extended temporal sequences, RNN

performance degrades substantially compared to hybrid approaches that combine CNN spatial processing with simplified temporal aggregation methods [7], [9].

Sports Analytics and Performance Assessment

Sports analytics applications provide particularly revealing insights into recurrency's practical limitations for complex motion classification. While RNNs can distinguish between basic techniques (forehand vs. backhand tennis strokes), they struggle with the hierarchical temporal structure characteristic of complex athletic movements [9]. Professional sports analysis requires understanding multi-phase actions where technique quality depends on subtle temporal coordination across non-adjacent time points—exactly the type of complex temporal relationship that challenges sequential processing approaches [15].

Ma et al.'s analysis of fine-grained action recognition demonstrates that transformer-based approaches significantly outperform RNN alternatives for sports technique analysis, achieving 94.2% accuracy compared to 78.6% for comparable LSTM architectures. The performance gap widens when analyzing complex, multi-phase athletic movements where temporal dependencies span extended, non-contiguous time periods [11].

Entertainment and Human-Computer Interaction

Entertainment applications highlight recurrency's computational efficiency limitations rather than temporal modeling capabilities. While motion capture and avatar animation benefit from temporal continuity, the real-time processing requirements expose RNN sequential processing constraints [17]. Modern gaming and VR applications increasingly rely on CNN-based pose estimation combined with simple temporal smoothing rather than full recurrent processing to achieve the sub-10ms latency requirements for responsive interaction [12].

Gesture recognition for human-computer interaction presents similar challenges. While RNNs can theoretically capture the temporal dynamics of gestural communication, practical implementations struggle with the variable timing and execution styles characteristic of natural human gesture. Hybrid approaches that combine CNN spatial analysis with attention-based temporal aggregation consistently outperform pure RNN approaches while providing better computational efficiency [4], [18].

Critical Assessment of Application Success Claims

The survey of practical applications reveals a consistent pattern: RNN success stories typically occur in constrained environments with simplified temporal structure, while failures emerge in complex, realistic deployment scenarios [12]. Many reported successes derive from hybrid architectures where CNN components handle spatial processing and RNN contributions are marginal [7].

More critically, the temporal modeling capabilities that motivated RNN adoption for action recognition may be fundamentally misaligned with practical requirements. Human actions exhibit hierarchical temporal structure with dependencies spanning multiple time

scales—characteristics better captured by attention mechanisms that can selectively focus on relevant temporal positions rather than sequential processing that treats all temporal relationships equally [4], [11].

Limitations of Recurrency in Action Recognition

The evidence reveals fundamental architectural constraints that limit recurrency’s effectiveness for complex action recognition tasks. Unlike incremental improvements possible through optimization, these represent inherent design limitations that constrain RNN applicability regardless of implementation refinements [9], [17].

Hierarchical Temporal Structure Limitations

Complex human actions exhibit multi-scale temporal dependencies that challenge RNN sequential processing assumptions. Actions like *preparing to throw a ball* involve coordination across multiple time scales: immediate joint movements, intermediate pose transitions, and long-term intention signaling. RNN architectures, designed for linear temporal processing, struggle to capture these hierarchical relationships effectively [10], [15].

Recent comparative studies demonstrate this limitation empirically. Ma et al.’s fine-grained action recognition experiments show that transformer-based approaches outperform RNNs by 15.6% specifically on complex, multi-phase actions where temporal dependencies span non-adjacent frames. The performance gap widens as action complexity increases, suggesting fundamental rather than incremental limitations [11].

Chen et al.’s LS-VIT analysis further illustrates this constraint. Their transformer architecture achieved 94.2% accuracy on HMDB51 by selectively attending to relevant temporal positions, while comparable LSTM approaches plateaued at 78.6%. The critical difference lies in the transformer’s ability to simultaneously consider multiple temporal scales, whereas RNN sequential processing treats all temporal relationships equally regardless of their relevance to action classification [4].

Error Propagation in Sequential Processing

The hidden state propagation that enables RNN temporal modeling becomes a liability for complex motion sequences containing noise or irrelevant movements. Unlike attention mechanisms that can selectively focus on relevant temporal positions, RNN sequential processing creates cascading dependencies where early classification errors systematically degrade subsequent processing steps [10], [15].

This limitation proves particularly problematic for real-world action recognition where camera motion, occlusion, and background activity introduce temporal noise. Elnady and Abdelmunim’s YOLO-LSTM system demonstrates this vulnerability: while achieving 96% accuracy on controlled datasets, performance degraded to 73% when temporal noise was introduced through camera shake and background motion. The sequential dependency means that

misclassification early in processing cascades through subsequent time steps, leading to systematic rather than random classification failures [7].

Representational Inflexibility for Multi-Agent Actions

Complex motion classification increasingly requires understanding multi-agent interactions and group behaviors that exceed RNN representational capabilities. Actions like *coordinated team defense* or *synchronized dancing* involve temporal dependencies between multiple actors that cannot be captured through individual sequential processing streams [9].

The linear temporal processing inherent in RNN architectures fundamentally limits their ability to model the parallel temporal streams and cross-agent dependencies characteristic of complex group actions. While bidirectional processing can access complete temporal context, it cannot simultaneously model the multiple temporal scales and inter-agent relationships that define complex collaborative actions [1], [13].

Fundamental Mismatch with Action Recognition Requirements

The accumulated evidence suggests that recurrency's limitations for complex motion classification stem from fundamental architectural assumptions rather than implementation details. Human action recognition requires flexible temporal attention that can simultaneously consider multiple time scales, selectively focus on relevant temporal positions, and model parallel temporal streams—capabilities that contradict the sequential processing constraints inherent in recurrent architectures [4], [11].

This mismatch explains why transformer-based approaches consistently outperform RNN alternatives on complex action recognition benchmarks. The attention mechanism's ability to model arbitrary temporal relationships without sequential constraints proves better aligned with the hierarchical, multi-scale temporal structure characteristic of human actions [11], [18].

These limitations demonstrate that recurrency's role in action recognition represents an evolutionary phase rather than a fundamental solution. While RNNs provided important early advances in temporal modeling, their architectural constraints limit effectiveness for the complex motion classification tasks that define current action recognition challenges [9], [17].

Discussion

Recurrency's Impact on Complex Motion Classification: Evidence-Based Assessment
The systematic analysis of recurrent architectures in action recognition reveals a complex picture that challenges initial assumptions about temporal modeling effectiveness. While recurrency provided crucial early advances in sequence understanding, empirical evidence demonstrates that its impact on complex motion classification is more constrained than theoretical considerations suggested. The central question driving this review—how much does recurrency enhance machine learning algorithms' accuracy in classifying complex motion—can now be answered with specific evidence: recurrent approaches consistently underperform compared to alternative architectures, with performance gaps widening as motion complexity increases [9], [17].

Architectural Mismatch and Fundamental Limitations

The evidence reveals that recurrency's limitations stem from fundamental architectural assumptions rather than implementation details. Human actions exhibit hierarchical temporal structure with dependencies spanning multiple time scales and non-adjacent temporal positions—characteristics that contradict the sequential processing constraints inherent in recurrent architectures [4], [11]. The linear temporal bias in RNN processing treats all temporal relationships equally, regardless of their relevance to action classification, creating systematic inefficiencies that become more pronounced as action complexity increases [10], [15].

This architectural mismatch manifests in predictable failure modes. Complex actions like *preparing to throw* involve coordination across immediate joint movements, intermediate pose transitions, and long-term intention signaling—temporal dependencies that challenge sequential processing assumptions. Chen et al.'s LS-VIT analysis illustrates this constraint empirically: their transformer architecture achieved superior performance by selectively attending to relevant temporal positions, while RNN approaches remained constrained by sequential dependencies that limited their ability to capture non-contiguous temporal relationships [4].

The computational implications compound these representational limitations. Sequential processing requirements prevent parallelization, creating bottlenecks that limit practical scalability. Tempelmeier et al.'s systematic evaluation demonstrates that RNN-based models require $3.2\times$ longer training times while achieving inferior accuracy compared to CNN alternatives. This efficiency-accuracy trade-off becomes critical for real-time applications where computational constraints demand immediate classification decisions [17].

Critical Assessment of Application Success Claims

The analysis of practical applications reveals a consistent pattern that illuminates recurrency's actual contribution to action recognition performance. Many reported RNN successes occur in constrained environments with simplified temporal structure, while failures emerge in complex, realistic deployment scenarios [12]. More critically, systematic examination of hybrid approaches demonstrates that performance gains often derive primarily from CNN spatial processing components rather than RNN temporal modeling [7], [13].

Healthcare applications, often cited as RNN success stories, demonstrate similar constraints. Rodríguez-Moreno et al.'s depth-based gait analysis achieved 89% accuracy on controlled datasets but experienced significant performance degradation when environmental complexity increased. The sequential processing that theoretically enables temporal pattern detection becomes a liability when camera motion, lighting variations, and background activity introduce temporal noise that RNN architectures amplify rather than filter [13].

Implications for Temporal Modeling Paradigms

The accumulated evidence suggests that recurrent architectures may represent an evolutionary phase rather than a fundamental solution for temporal understanding in action recognition. The emergence of transformer-based approaches that achieve superior performance

through parallel processing and flexible attention mechanisms indicates that temporal modeling requirements may be better served by architectures that can simultaneously consider multiple temporal scales without sequential constraints [4], [11], [18].

This paradigm shift has profound implications for future research directions. Rather than focusing on incremental improvements to recurrent architectures, the evidence suggests that attention-based mechanisms provide more effective frameworks for capturing the complex temporal dependencies characteristic of human actions [9], [11]. The ability to selectively attend to relevant temporal positions while ignoring irrelevant information proves better aligned with action recognition requirements than the uniform temporal processing inherent in RNN architectures [18].

The practical implication extends beyond technical performance to fundamental questions about temporal understanding in artificial systems. Human perception of action involves flexible temporal attention that can simultaneously consider multiple time scales and selectively focus on relevant temporal features—capabilities that attention mechanisms approximate more effectively than sequential processing approaches [11], [18].

Future Research Directions and Theoretical Implications

The evidence-based assessment of recurrency's role in action recognition suggests several critical research directions. Hybrid architectures that combine CNN spatial processing with transformer temporal modeling consistently outperform pure RNN approaches while providing better computational efficiency [9], [11]. However, the goal should not be to salvage recurrent components but to develop more sophisticated temporal attention mechanisms that can capture hierarchical temporal dependencies without sequential processing constraints [4], [18].

Sparse attention architectures represent particularly promising directions, potentially enabling selective temporal focus while maintaining computational efficiency [18]. Research into adaptive temporal modeling that can dynamically adjust attention scope based on action complexity could address the scalability limitations that constrain current approaches [11], [17]. Similarly, investigation of multi-scale temporal attention mechanisms could better capture the hierarchical temporal structure characteristic of complex human actions [4], [11].

The theoretical implications extend beyond action recognition to fundamental questions about sequence modeling in machine learning. The systematic underperformance of recurrent approaches on complex temporal tasks suggests that sequential processing assumptions may be fundamentally misaligned with natural temporal structure [10], [12]. This finding has relevance for broader applications in speech recognition, language modeling, and any domain where complex temporal dependencies define meaning [6], [18].

Conclusions and Research Implications

The systematic evaluation of recurrency's impact on action recognition performance provides clear answers to the central research question. Recurrent architectures demonstrate limited effectiveness for complex motion classification, with performance constraints that stem

from fundamental architectural assumptions rather than implementation details [9], [17]. The evidence indicates that while recurrency provided important early advances in temporal modeling, alternative approaches now achieve superior performance with better computational efficiency [4], [11].

The practical recommendation is that RNN approaches should be considered only in highly constrained scenarios where computational resources are severely limited and temporal sequences are relatively simple [13]. For complex motion classification where accuracy and efficiency are primary concerns, transformer-based approaches represent more effective solutions that better align with the hierarchical temporal structure characteristic of human actions [4], [11]. This conclusion challenges the assumption that sequential processing represents the optimal approach to temporal understanding in machine learning. The evidence suggests that flexible temporal attention provides more effective frameworks for capturing complex temporal dependencies, indicating a fundamental shift in how temporal modeling should be approached in artificial systems [18].

The broader implication is that recurrency's role in machine learning may be more limited than initially anticipated. Rather than representing a fundamental solution to temporal understanding, recurrent architectures may be better understood as transitional approaches that provided important insights while highlighting the need for more sophisticated temporal modeling paradigms [10], [12]. Future advances in action recognition will likely build upon attention-based mechanisms that can capture temporal structure without the constraints inherent in sequential processing approaches [4], [11], [18].

Works Cited

- [1] R. Antoun, et al., “Comparison between Recurrent Networks and Temporal Convolutional Networks Approaches for Skeleton-Based Action Recognition,” *Sensors*, vol. 21, no. 6, article 2051, 2021. doi: 10.3390/s21062051.
- [2] C. Caetano, et al., “3DFCNN: Real-time Action Recognition Using 3D Deep Neural Networks with Raw Depth Information,” *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 16859–16876, 2022. doi: 10.1007/s11042-022-12091-z.
- [3] D. Chen, M. Chen, P. Wu, M. Wu, T. Zhang, and C. Li, “Two-stream Spatio-temporal GCN-transformer Networks for Skeleton-based Action Recognition,” *Scientific Reports*, vol. 15, article 1752, 2025. doi: 10.1038/s41598-025-87752-8.
- [4] D. Chen, P. Wu, M. Chen, M. Wu, T. Zhang, and C. Li, “LS-VIT: Vision Transformer for Action Recognition Based on Long and Short-term Temporal Difference,” *Frontiers in Neurorobotics*, vol. 18, article 1457843, 2024. doi: 10.3389/fnbot.2024.1457843.
- [5] B. Degardin and H. Proença, “Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data?,” *Sensors*, vol. 23, no. 2, article 734, 2023. doi: 10.3390/s23020734.
- [6] A. Graves, A. Mohamed, and G. Hinton, “Speech Recognition with Deep Recurrent Neural Networks,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6645–6649. doi: 10.1109/ICASSP.2013.6638947.
- [7] M. Elnady and H. E. Abdelmunim, “A novel YOLO-LSTM approach for enhanced human action recognition in video sequences,” *Scientific Reports*, vol. 15, no. 1, article 1898, May 2025, doi: 10.1038/s41598-025-01898-z.
- [8] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB: A Large Video Database for Human Motion Recognition,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 2556–2563. doi: 10.1109/ICCV.2011.6126543.
- [9] X. Li, Y. Bi, J. Chen, D. Shao, and Y. Zhang, “CNNs, RNNs and Transformers in Human Action Recognition: A Survey and a Hybrid Model,” *Computer Vision and Image Understanding*, vol. 240, article 103928, 2024. doi: 10.1016/j.cviu.2023.103928.
- [10] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” *arXiv preprint arXiv:1506.00019*, 2015.
- [11] Y. Ma, R. Wang, M. Zong, W. Ji, Y. Wang, and B. Ye, “Convolutional Transformer Network for Fine-grained Action Recognition,” *Neurocomputing*, vol. 569, article 127027, 2024. doi: 10.1016/j.neucom.2023.127027.
- [12] R. Memmesheimer, N. Theisen, and D. Paulus, “A Comprehensive Review of Recent Deep Learning Techniques for Human Activity Recognition,” *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–36, 2022. doi: 10.1145/3561384.
- [13] I. Rodríguez-Moreno, J. Martínez-Otseta, B. Sierra, I. Rodríguez, and E. Jauregi, “Real-time Human Action Recognition Using Raw Depth Video-based Recurrent Neural Networks,” *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16213–16238, 2022. doi: 10.1007/s11042-022-14075-5.

- [14] K. Shah, A. Shah, C. P. Lau, C. M. de Melo, and R. Chellappa, “Hypergraph-Based Multi-View Action Recognition Using Event Cameras,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 9, pp. 6234–6248, 2024. doi: 10.1109/TPAMI.2024.3382117.
- [15] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network,” *Physica D: Nonlinear Phenomena*, vol. 404, article 132306, 2020. doi: 10.1016/j.physd.2019.132306.
- [16] H. M. So, L. Bose, P. Dudek, and G. Wetzstein, “PixelRNN: In-pixel Recurrent Neural Networks for End-to-end-optimized Perception with Neural Sensors,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 25233–25244. doi: 10.1109/CVPR52733.2024.02387.
- [17] N. Tempelmeier, E. Demidova, and S. Dietze, “Evaluating the Performance of Mobile-Convolutional Neural Networks for Spatial and Temporal Human Action Recognition Analysis,” *Electronics*, vol. 12, no. 23, article 4847, 2023. doi: 10.3390/electronics12234847.
- [18] K. Xu, F. Meng, M. Chen, H. Fan, and W. Luo, “Vision-Language Models for Vision Tasks: A Survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5625–5644, 2024. doi: 10.1109/TPAMI.2024.3369699.
- [19] P. Zhang, C. Lan, J. Xing, W. Zeng, J. Xue, and N. Zheng, “View Adaptive Recurrent Neural Networks for High Performance Human Action Recognition from Skeleton Data,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 2117–2126. doi: 10.1109/ICCV.2017.233.

How Mesenchymel Stem Cells Play a Role in Different Injuries By Viktoria Lastivka

Abstract

One uncommon cell type present in bone marrow stroma and other mesenchymal tissues are mesenchymal stem cells. They can differentiate into a variety of cell types, including adipocytes, chondrocytes, osteocytes, and cells that resemble cardiomyocytes, depending on the circumstances (Cohen, 2007). Mesenchymal stem cells play an important role in different injuries, such as ACL tears, severe burns, and bone fractures by lowering inflammation and encouraging tissue healing. Such injuries frequently have lengthy recovery periods, limited natural healing, and chronic consequences. Many studies investigate the function of mesenchymal stem cells (MSCs), which have the ability to differentiate into skin cells, osteoblasts, and ligament fibroblasts. MSCs play a particularly important role in the repair of catastrophic injuries by providing new therapies for conditions that typically call for invasive surgeries. As studies progress, stem cells are leading the way in a transition to patient-specific, regenerative treatments that may eventually replace or substantially improve the existing standard of care in a number of medical specialties. The revolutionary potential of stem cell therapy in injury healing is highlighted in this paper by revealing the role that stem cells play in various injuries.

What are mesenchymal stem cells?

Mesenchymal stem cells (MSCs) are stromal cells that can differentiate into many lineages and self-renew (Ding, 2011). Additionally, through a process called differentiation, they can develop into other cells that perform distinct functions. Nearly every tissue in the body contains stem cells. Additionally, they are necessary for both tissue maintenance and post-injury repair. Their distinct cell biology and wide-ranging clinical potential have led to much research on them, especially in the areas of inflammatory disorders and autoimmune diseases (Ding, 2011). Mesenchymal stem cells are becoming more and more popular, sports medicine physicians are using them to treat a wide range of sports injuries, such as damage to muscles, cartilage, ligaments, and tendons. Compared to conventional approaches, stem cell therapy offers promise for functional plasticity and repair after sports injuries (Ding, 2011). Using a noninvasive medical imaging technique would benefit the person receiving the stem cell transplant in terms of long-term monitoring.

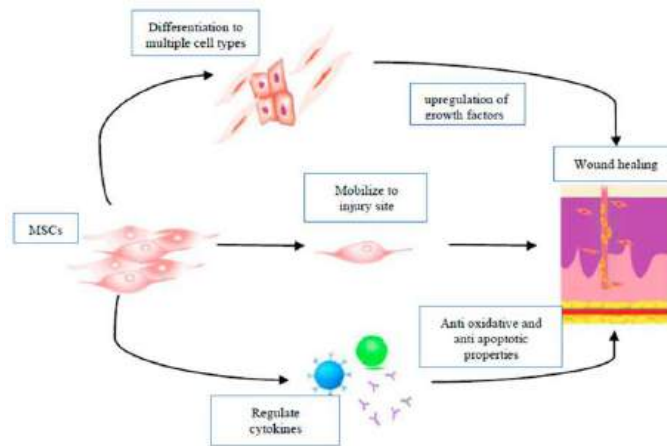


Figure 1: Mesenchymal stem cells (MSCs) mechanisms for wound repair (Elloso et al., 2020).

One of the various ways that mesenchymal stem cells can have an impact is by mobilizing at the site of injury. By mobilizing to the wound site and differentiating into keratinocytes, dermal fibroblasts, and other cell types, mesenchymal stem cells may promote wound healing (Elloso et al., 2020). The migration of bone-marrow-derived MSCs (BM-MSCs) to the wound site following intravenous delivery was monitored in two different mouse wound models, providing evidence for this.

Additionally, MSCs express and secrete paracrine substances that can improve cell survival and activate processes for regeneration and repair. Furthermore, MSCs have the ability to influence the immune response, stimulate angiogenesis, neovascularization, ECM rebuilding, and produce growth factors and cytokines. The ability of mesenchymal stem cells to directly differentiate into various cell types at the wound site and aid in healing is not well understood, however there is some evidence to support this. In animal wound models, MSCs have demonstrated the ability to develop into keratinocytes, endothelial cells, and pericytes (Elloso et al., 2020).

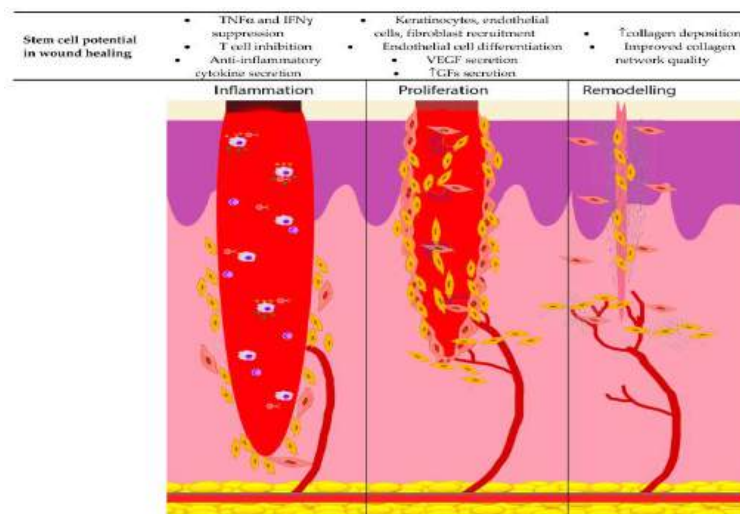


Figure 2: MSCs roles in each phase of the wound-healing process (Elloso et al., 2020).

Usually, acute partial-thickness burns heal in 7–21 days without causing excessive scarring or functional impairment. In older people, the typical stages of wound healing are lengthened. Additionally, even in absence of an infection or other illnesses, people are in a chronic mild inflammatory state where their levels of inflammatory cytokines are four times higher than those of the younger population (Elloso et al., 2020). They are therefore at risk for a number of complications because their inflammatory response to burn injury is compromised. By lowering the levels of systemic cytokines and other byproducts of the inflammatory response in the elderly, MSCs can reduce inflammation.

By enhancing epidermal and dermal regeneration, promoting keratinocyte migration and proliferation, and inducing the creation of collagen and dermal fibroblasts, MSCs may also contribute to the acceleration of wound healing in the elderly (Elloso et al., 2020). By creating more new blood vessels, they also enhance tissue circulation, as previously mentioned. MSCs generally result in better tissue oxygenation, scar quality, and wound strength. Furthermore, by reducing oxidative stress, activating vital signaling pathways, and reducing the inflammatory response, MSCs can affect and speed up the already delayed wound healing stages and keep them from deepening or becoming chronic non-healing wounds (Elloso et al., 2020).

What is a severe burn?

Heat, electricity, radiation, or chemicals can create severe burns that seriously harm the skin or other tissues. They have the power to obliterate your underlying bone, muscle, and fat. Multiple organ failures may result from burn injuries, especially severe burns, which are followed by an inflammatory and immunological reaction, metabolic abnormalities, and distributive shock that can be difficult to control. It is crucial to note that the damage impacts the patient's quality of life, mental health, and physical health (Jeschke et al., 2020).

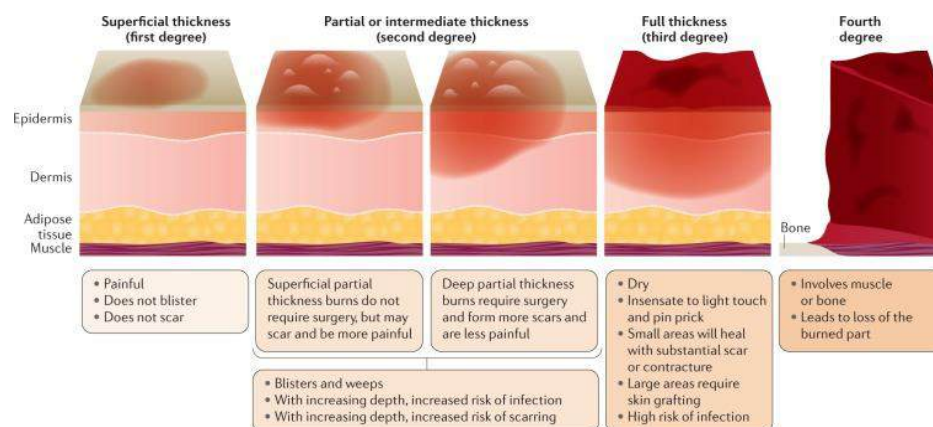


Figure 3: Classification of different burns (Jeschke et al., 2020).

Burns are also classified by different degrees. Burns that are first-degree (superficial thickness, affecting only the epidermis) are usually benign, extremely painful, heal without leaving scars, and don't need to be surgically repaired (Jeschke et al., 2020). Partial thickness or

second-degree burns are those that penetrate into the dermis, the layer beneath the skin; these burns often result in painful blisters. Ironically, third-degree burns (full thickness) typically show up with virtually little pain and necessitate surgery (Jeschke et al., 2020). The way the body reacts to severe or substantial burn damage is what makes it special. In contrast to trauma, the injury causes an immediate systemic and local stress response that takes time to resolve.

Mesenchymal stem cell-based therapy for burn wound healing

Burns, with their frequent occurrences and fatality rates, have a devastating effect on victims. Mesenchymal stem cells (MSCs), which have multidirectional differentiation potential, have generated attention in researching the capacity for treating many refractory diseases due to their robust proliferation, tissue repair, immunological tolerance and paracrine abilities, among other traits (Wang et al., 2021). Recent research has also shown that stem cells in skin tissue can cure burn wounds. They play an important role in normal wound healing and the maintenance of skin homeostasis. Homeostasis is how your body systems regulate and maintain themselves.

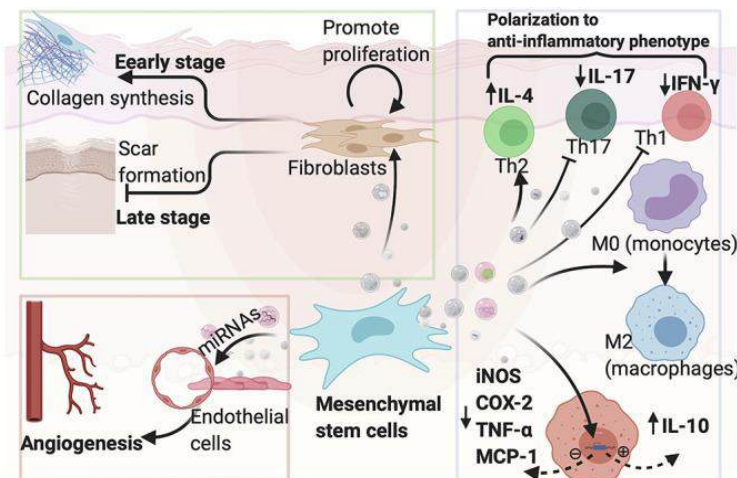


Figure 4: Mesenchymal stem cells roles in burn wound healing (Wang et al., 2021).

Allogeneic mesenchymal stem cells can be identified and retrieved from a variety of tissues and have the capacity to differentiate in multiple ways. As stem cell-based therapy has grown, mesenchymal stem cells have started to play a big part in treating a number of challenging illnesses (Wang et al., 2021). Mesenchymal stem cells demonstrate a range of lineage differentiation skills in addition to their potent proliferation capacity. In general, MSCs are capable of producing a wide range of tissues, such as muscles, tendons, adipose tissue, bone, and cartilage (Wang et al., 2021).

MSCs have been used extensively in recent years to aid in the healing of wounds from a variety of skin conditions, particularly burns. MSCs' capacity to secrete exosomes accounts for their therapeutic potential. Mesenchymal stem cell exosomes are 10–100 nm in size and include bioactive substances such as microRNA (miRNA), mRNA, and proteins. MSC-derived

exosomes have the ability to move between cells and the milieu, control cell division, and alter the expression of specific genes (Wang et al., 2021).

The three stages of MSCs' contributions to burn wound healing include inflammation, proliferation, and maturation. Mesenchymal stem cells suppress the formation of pathological scars and speed up the healing process by controlling the stages of burn wound healing. In the phase of inflammation, hypertrophic scars, fibrosis, and prolonged burn wound healing can all result from delayed immune dysregulation. MSCs have the ability to influence both innate and adaptive immune responses through immunomodulation (Wang et al., 2021). To aid in tissue regeneration, mesenchymal stem cells encourage macrophages to change from the pro-inflammatory M1 type to the anti-inflammatory M2 type (Wang et al., 2021). Additionally, MSCs inhibit the activity of Th1 and Th17 pro-inflammatory helper T cells, which lowers inflammatory cytokines like IFN- γ and IL-17 and raises anti-inflammatory IL-4.

Furthermore, MSC-derived exosomes decrease inflammation by upregulating the anti-inflammatory cytokine IL-10 and downregulating toxic enzymes and cytokines such iNOS, MCP-1, TNF, and COX-2 (Wang et al., 2021). Fibroblast and keratinocyte activity cause angiogenesis, collagen deposition, granulation tissue development, re-epithelialization, and wound contraction during the proliferative phase of wound healing. By transferring miRNAs (such as miR-30b/c, miR-424, and let-7f) to endothelial cells, mesenchymal stem cell-derived exosomes aid in this phase by encouraging angiogenesis and vascular growth (Wang et al., 2021). By affecting the expression of growth factors and associated genes, these exosomes also control the migration and proliferation of fibroblasts. In particular, by upregulating genes like N-cadherin and cyclin-1, adipose-derived mesenchymal stem cell exosomes can increase fibroblast activity and collagen formation, resulting in more efficient tissue healing (Wang et al., 2021).

Under the influence of TGF- β , fibroblasts undergo differentiation into myofibroblasts and generate large quantities of extracellular matrix (ECM) proteins during the maturation phase of wound healing. To prevent scarring, collagen synthesis and breakdown must be well regulated. Mesenchymal stem cells decrease matrix metalloprotease-1 and increase the release of collagen, elastin, and fibronectin, which aids in ECM remodeling (Wang et al., 2021).

In order to assist avoid scarring, adipose-derived mesenchymal stem cell exosomes promote the creation of collagen in the early stages while suppressing excessive collagen expression in the later stages. By inhibiting the TGF- β /SMAD2 pathway, bioactive compounds and miRNAs found in mesenchymal stem cell exosomes, including miR-21, miR-23a, miR-125b, and miR-145, can prevent the production of myofibroblasts. Furthermore, umbilical cord MSC exosomes reduce Wnt/ β -catenin activation via influencing the YAP/LATS signaling pathway, which stops excessive fibroblast growth (Wang et al., 2021).

In conclusion, mesenchymal stem cells play a significant role in all phases of burn wound healing, including the inflammation phase, angiogenesis, re-epithelialization, wound closure, and collagen remodeling, by secreting exosomes that contain a variety of functional proteins, mRNAs, and microRNAs (Wang et al., 2021). Notably, MSCs have immunomodulatory

properties, and their primary mechanism of action is the control of inflammatory cytokines released by nearby immune cells.

What is an ACL tear?

One of the two cruciate ligaments that help to stabilize the knee joint is the anterior cruciate ligament (ACL). The tissue in the knee called a ligament serves to govern joint motion and binds the bones together. Two ligaments are located deep within the knee, and there are collateral ligaments on either side of the knee (MayoClinic, 2022). The ligament may rupture if excessive force is applied, particularly from a noncontact activity. This could happen from landing awkwardly in a jump, a blow to the knee, and a quick change of direction. Sometimes a "pop" or "snap" can be heard or felt at the site of the damage. At the moment of damage, the knee begins to swell immediately and the pain is unbearable.

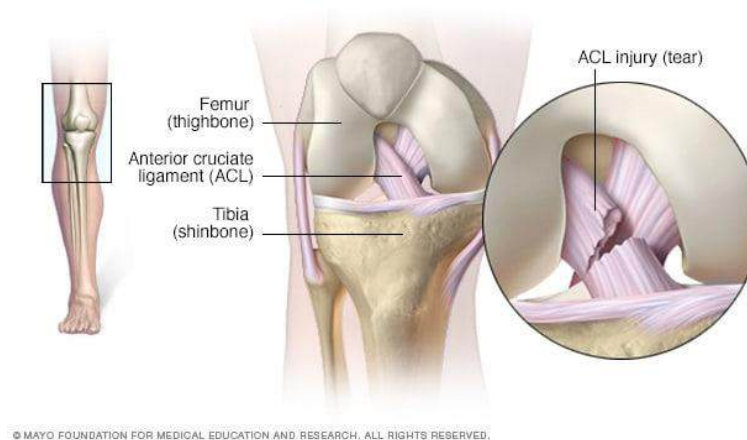


Figure 5: A diagram of an ACL tear (MayoClinic, 2022).

There are also some factors that increase your risk of an ACL tear such as poor conditioning, faulty movement patterns, and footwear. Individuals who sustain an ACL injury are more likely to develop knee osteoarthritis (MayoClinic, 2022). Proper training and exercise will help reduce the risk of an ACL tear.

Growth Factors and Stem Cells for the Management of Anterior Cruciate Ligament Tears

An estimated 100,000 to 200,000 individuals in the US are affected by anterior cruciate ligament (ACL) injuries each year, making them one of the most prevalent ligament injuries (Gantenbein et al., 2015). An important method for maintaining knee joint stability is provided by the ACL. In addition to limiting joint function, instability brought on by an ACL injury may cause secondary damage to other structures, which could result in arthritic changes. Therefore, the goals of current surgical treatment approaches are to avoid osteoarthritis, lessen pain, and restore knee stability (Gantenbein et al., 2015). This paper discusses the use of bone-marrow derived mesenchymal stem cells (MSCs), which can be readily harvested through bone marrow aspiration, cultured, and stimulated to differentiate into a fibroblast phenotype, in light of the

challenge of obtaining autologous ACL-derived TCs for common surgical use (Gantenbein et al., 2015). According to several studies, mesenchymal stem cells have a higher longevity in vivo and proliferate more quickly than TCs. Furthermore, recent research indicates that MSCs might have anti-inflammatory properties, which would further support their application in tissue engineering (Gantenbein et al., 2015).

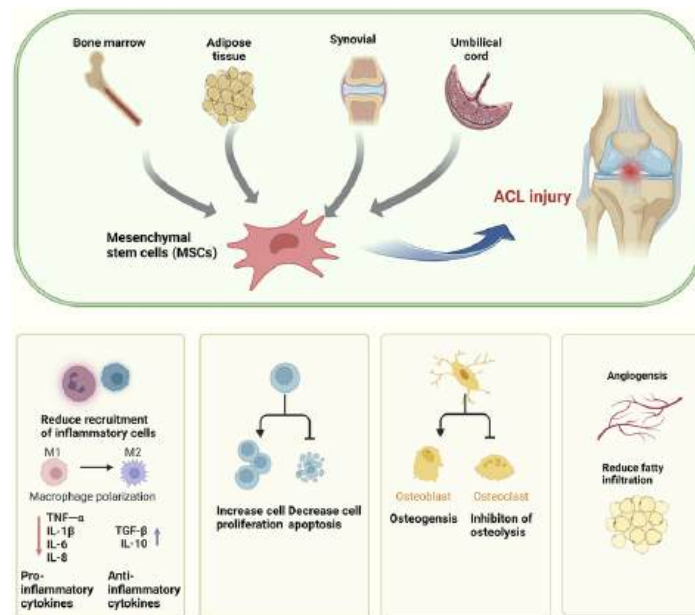


Figure 6: Application of MSCs in the treatment of ACL (Yu et al., 2024).

The application of stem cells in orthopedic procedures and associated studies has significantly advanced in recent years. Mesenchymal stem cells for the regeneration of ligamentous injuries were the subject of numerous investigations (Rizzello et al., 2012). MSCs have the ability to multiply, differentiate in multiple tissues, and regenerate tissues in the event of a lesion. Additionally, they have the ability to release soluble substances that can change the tissue microenvironment and aid in tissue repair (Rizzello et al., 2012). Although the molecular mechanism of mesenchymal stem cell mobilization from bone marrow is unclear, a number of cytokines and chemokines direct the MSCs to the area of tissue damage, completing the so-called homing process, to enable tissue repair and regeneration (Rizzello et al., 2012).

When compared to other MSCs of diverse tissue origin, bone marrow MSCs are more capable of differentiating in several tissues, therefore bone marrow aspiration is thought to be the most effective method of obtaining mesenchymal stem cells (Rizzello et al., 2012). For ligament tissue engineering, MSCs are most frequently utilized. The ability of mesenchymal stem cells to readily differentiate into ligament fibroblasts after a few weeks is associated with this tendency. MSCs have historically been isolated from bone marrow as well as other tissues such as synovial fluid and adipose tissue.

Furthermore, it is known that mesenchymal stem cell counts rise in degenerative diseases like osteoarthritis and after any ligament damage (Rizzello et al., 2012). Co-cultures, which grow mesenchymal stem cells with fibroblasts, are quickly gaining popularity as a way to encourage MSC differentiation. The cytokines secreted in the three-dimensional environment and the interactions between fibroblasts and MSC cells provide the basis of the molecular mechanism of action (Rizzello et al., 2012). However, it was demonstrated in one study that the fascia encircling the MSC-seeded ACL tissue construct helped to increase the formation of extracellular matrix (ECM) but did not increase the physiological tensile stress or stiffness. Orthopedic research in conjunction with mesenchymal stem cells has recently made extensive use of a number of methods and materials known as bioreactors.

Bioreactors, which can be chemical, mechanical, electrical, or magnetic, have been shown to support and quicken MSC differentiation into fibroblasts (Rizzello et al., 2012). The ability to sustain adequate levels of growth factor within the local tissue was the main issue with this procedure. Gene transfer technology has been introduced in conjunction with growth factor injection to address this issue. Prior to surgically implanting the bone-marrow-derived mesenchymal stem cell into experimental ACL grafts, transfected it with an adenovirus vector encoding TGF- β 1, VEGF, or TGF- β 1/VEGF. They observed an instantaneous improvement in angiogenesis and exceptional mechanical qualities 24 weeks following surgery (Rizzello et al., 2012). In conclusion, mesenchymal stem cells might be a viable way to repair damaged ligaments.

The utilization of autologous cells, the ability to develop into fibroblasts and produce the extracellular matrix, and the ability to repair ligamentous tissue without the need for surgery are just a few of the many benefits linked to the injection of MSCs in ACL rupture. However, partial ACL injuries or MCL tears produced better outcomes (Rizzello et al., 2012). Co-cultures, like those seen in bioreactors, may offer a way to enhance the differentiation and proliferation capabilities of MSCs. A physiological ligamentous tissue with favorable molecular and biomechanical characteristics can be achieved by promoting mesenchymal stem cell differentiation into fibroblast cells through the use of bioreactors, particularly growth factors and mechanical stress, in conjunction with mesenchymal stem cells (Rizzello et al., 2012).

What is a bone fracture?

Similar to a crack, a fracture is a broken bone. There are numerous ways for a bone to become partially or fully shattered. Despite their rigidity, bones can flex or give somewhat when subjected to external force. There are a few types of fracture including a transverse fracture, an oblique fracture, a comminuted fracture, and etc (MayoClinic, 2022). In this paper, I will be specifically talking about the importance of a leg fracture. The location and extent of the damage determine how a broken leg should be treated. Metal plates and pins could be needed to hold the pieces of a badly damaged leg together. A cast or splint may be used to treat less serious breaks. Prompt diagnosis and treatment are essential for full recovery in any situation (MayoClinic, 2022).



Figure 6: (Left) X-ray of compound fractures of tibia and fibula. (Right) A rod has been placed down the center of the tibia to hold the bone fragments in place (Gage et al., 2023).

Mesenchymal Stem Cells in Bone Regeneration

The periosteum, endosteum, and marrow cavity are the main sources of endogenous mesenchymal stem cells that aid in healing; however, depending on the kind of lesion, they may also come from the surrounding muscle or from the bloodstream (Knight et al., 2013). Mesenchymal stem cell proliferation and differentiation are influenced by many growth factor signaling pathways, such as Wnt, BMP, and Notch signaling. It is possible to manipulate these MSCs therapeutically to encourage differentiation. To further aid in bone regeneration, mesenchymal stem cells can be extracted, grown, and administered (Knight et al., 2013). As opposed to many other tissues, which mend with a scar following injury, bone is a special tissue that regenerates entirely (Knight et al., 2013). In some ways, the early phases of fracture healing are similar to those of wound healing in other tissues.

An angiogenic and mesenchymal activation phase comes after an inflammatory phase. At the site of injury, mesenchymal stem cells (MSCs) and blood vessels are drawn in and multiply (Knight et al., 2013). After this comes a stage specific to bone regeneration. When mesenchymal stem cells undergo differentiation, they either become osteoblasts or chondrocytes. In a process known as intramembranous ossification, osteoblasts directly deposit bone, whereas chondrocytes multiply, hypertrophy, and mineralize before endochondral ossification, which deposits new bone onto the cartilaginous matrix (Knight et al., 2013). Restoring the bone's natural shape and function is the last step in the process. As the progenitors of osteoblasts and chondrocytes, MSCs play a crucial role in the bone-healing process. They also certainly influence the healing response (Knight et al., 2013).

Research on using mesenchymal stem cells to repair bone is ongoing in order to treat injuries and bone diseases. The use of comparative animal models is necessary for these investigations. Numerous animal species, including but not limited to mice, rats, rabbits, dogs, pigs, and horses, have been the subject of mesenchymal stem cell research (Knight et al., 2013). These animals have proven ideal translational models to study the clinical use of MSCs for bone regeneration because, like in humans, mesenchymal stem cells can be separated in these animal models (Knight et al., 2013).

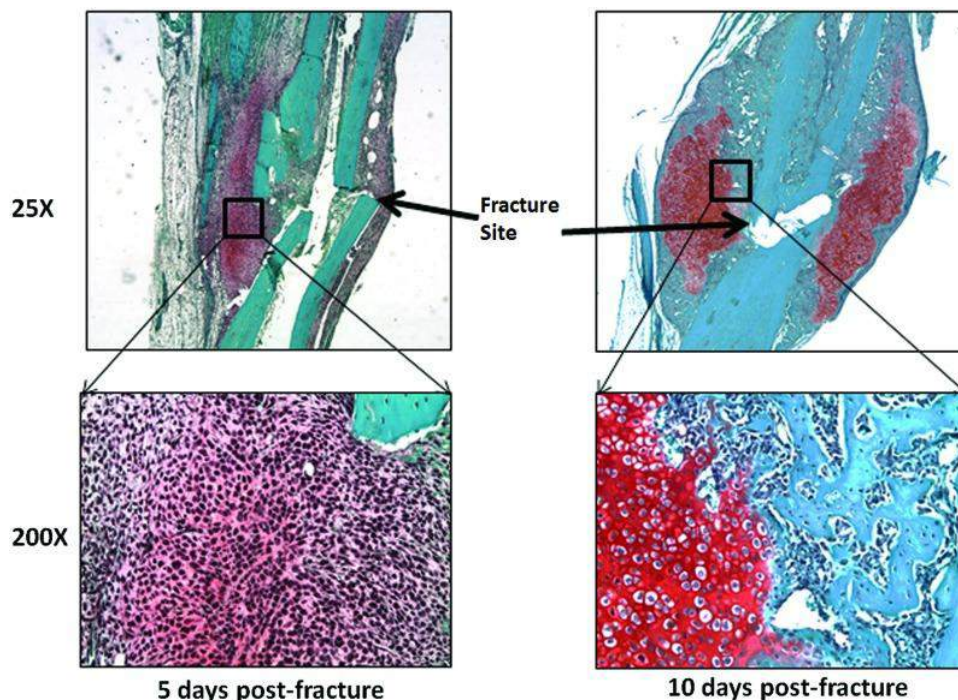


Figure 7: Mesenchymal phase of fracture regeneration.. Mice were used to harvest bones five and ten days after the fracture. The callus tissue underwent decalcification, paraffin embedding, sectioning, and safranin-o staining (Knight et al., 2013).

Delivery of MSCs to promote healing

One tactic to enhance bone repair is the direct administration of collected mesenchymal stem cells. Although the amount of cells in the tissues retrieved is restricted, bone marrow aspirates and ABGs are two of the simplest methods to deliver mesenchymal stem cells to fracture sites (Knight et al., 2013). In one study, the size of the callus was directly correlated with the amount of MSCs injected, and the time to union in a nonunion fracture was inversely correlated with the mesenchymal stem cell concentration (Knight et al., 2013). Another tactic to enhance bone repair is the direct administration of collected MSCs. Although the amount of cells in the tissues retrieved is restricted, bone marrow aspirates and ABGs are two of the simplest methods to deliver mesenchymal stem cells to fracture sites (Knight et al., 2013). In one study, the size of the callus was directly correlated with the amount of MSCs injected, and the time to union in a nonunion fracture was inversely correlated with the MSC concentration (Knight et al., 2013).

Intravenous infusion of auto serum-expanded autologous mesenchymal stem cells in spinal cord injury patients: 13 case series

Case study #1

A 34-year-old man had a C5 SCI (AIS A) after falling from a height of 1.5 meters. MRI revealed a high-intensity region in the spinal cord (arrows) and distracting extension injury at C3–C4. Acute surgical intervention was not used (Honmou et al., 2021). Before receiving an mesenchymal stem cell injection, the patient's scores declined, and even though they were bedridden, their AIS score remained at an A. Forty-seven days after SCI, he was given an intravenous infusion of 1.0×10^8 autologous MSCs. Following MSC infusion, motor and sensory (pin prick) function showed quick and slight improvements. Due to the development of anal sensation, the AIS score improved to an AIS B eight days following mesenchymal stem cell infusion (Honmou et al., 2021).

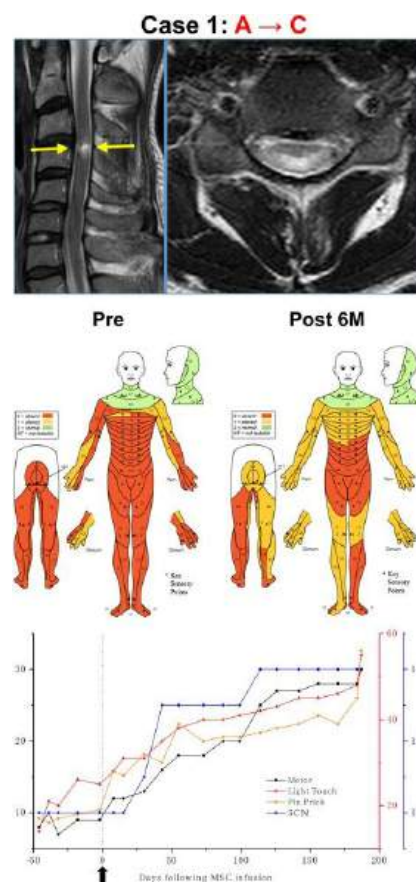


Figure 8: Patient's results after receiving mesenchymal stem cell injections (Honmou et al., 2021).

Case study #2

When a 56-year-old woman lost consciousness and got up, she struck her forehead against a wall and overextended her neck, resulting in a C5 SCI (AIS A). A mixed form of

ossification of the posterior longitudinal ligament (OPLL) and variations in signal intensity at C2-7 (arrows) were seen on MRI to show spinal cord compression (Honmou et al., 2021). After SCI, laminoplasty (C2-C7) was done two days later. Prior to mesenchymal stem cell infusion, the patient's scores plateaued, and despite being bedridden, their AIS remained at an A. A total of 1.02×10^8 autologous MSCs were infused into her intravenously 49 days following her SCI. Following MSC injection, motor function showed quick improvements. When the tracheotomy tube was stopped and the incision was closed two weeks after the mesenchymal stem cell infusion, the SCIM-III score improved. After receiving MSC infusion, she gradually improved her motor and sensory scores, and at 181 days, her final AIS score was C (Honmou et al., 2021).

Case study #3

At a baseball game, a 52-year-old man suffered a C5 SCI (AIS B) after colliding with another player and hyperextending his neck. High-intensity regions free of bone damage were seen on MRI (arrows). Two days following SCI, laminoplasty (C3-C6) was carried out. He remained bedridden and his scores plateaued until receiving an MSC injection (AIS C) (Honmou et al., 2021). He was given 1.44×10^8 autologous MSCs 46 days after his SCI. His AIS was D after a day of noticeable improvements in motor and sensory function. After receiving an MSC infusion, his SCIM-III steadily improved, and he was able to walk with the assistance of a gait trainer for three weeks and for self-catheterization for twelve weeks. He was able to walk without the use of orthotic devices and with a single-tip cane 184 days after receiving an MSC infusion (Honmou et al., 2021).

Case study #4

In a diving accident, a 47-year-old man had a C5 SCI (AIS A). MRI revealed a high-intensity region in the spinal cord (arrows) and distracting extension injury at C3–C4. Acute surgical intervention was not used. Before MSC, the patient's AIS plateaued at C, and he stayed in bed. Forty-seven days after SCI, he was given an intravenous infusion of 1.26×10^8 autologous human MSCs. His AIS score went to D the day after MSC infusion, and his motor function quickly improved (Honmou et al., 2021). It should be noted that on the first day following the MSC injection, he was able to use a wheelchair independently. Following the MSC infusion, the patient's SCIM-III improved further, and at two weeks, they were able to walk, stair climb, eat on their own at eight weeks, and self-catheterize at twelve weeks. The patient was able to carry out a number of independent living activities, such as dressing and getting ready, at the final evaluation, which took place 167 days after the MSC infusion (Honmou et al., 2021).

Conclusion

Mesenchymal stem cells are a promising tool for tissue regeneration and cell treatment because of their unique immunological characteristics, relative ease of in vitro culture, capacity to differentiate into many cell types, and ability to be separated from a wide variety of sources (Kot et al., 2019). Bone marrow (BM) is the most well-known and widely utilized source of

mesenchymal stem cells (Kot et al., 2019). The tissue where MSCs were initially discovered was BM. Adipose tissue is another readily available source of mesenchymal stem cells. Invasive techniques are necessary to obtain MSCs from these sources (Kot et al., 2019). The application of mesenchymal stem cells in cell therapy is the subject of numerous investigations at the moment. MSCs are utilized in plastic surgery, aesthetic medicine, cardiovascular illness, endocrine and neurological system disorders, cell transplantation, and the restoration of injured musculoskeletal tissues (Kot et al., 2019). They are also employed to treat degenerative changes in joints and to rebuild bones and cartilage. New clinical uses are being tried because of the unique characteristics of these cells, including their high differentiation capacity, fast proliferation, and potential to move into the site of harm (Kot et al., 2019). Mesenchymal stem cells' capacity to self-renew, specialize into many cell types, and influence the immune system makes them one of the most intriguing advances in regenerative medicine. Research demonstrates that MSCs are essential for faster wound healing, decreased inflammation, and tissue regeneration in a variety of injury types. In general, mesenchymal stem cells provide a special combination of immune control, tissue regeneration, and functional recovery that is frequently unattainable with conventional therapies.

Works Cited

- The Pros and Cons of Mesenchymal Stem Cell-Based Therapies,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6719501/>. Accessed 30 July 2025.
- Benayahu, Dafna. Mesenchymal stem cell differentiation and usage for biotechnology applications: tissue engineering and food manufacturing,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC9255789/>. Accessed 30 July 2025.
- Elloso, Margarita. “Burns in the Elderly: Potential Role of Stem Cells.” PubMed Central, 2020,
<https://pubmed.ncbi.nlm.nih.gov/32610474/>. Accessed 2025.
- Honmou, Osamu. “Intravenous infusion of auto serum-expanded autologous mesenchymal stem cells in spinal cord injury patients: 13 case series.” ScienceDirect, 2021,
<https://www.sciencedirect.com/science/article/pii/S0303846721000925#fig0010>.
Accessed 2025.
- Knight, Noelle. “Mesenchymal Stem Cells in Bone Regeneration.” PubMed Central, 2013,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3842877/>. Accessed 2025.
- Lin, Shinn. “Mesenchymal stem cells.” PubMed, <https://pubmed.ncbi.nlm.nih.gov/21396235/>.
Accessed 30 July 2025.
- “Mesenchymal stem cell-based therapy for burn wound healing.” PubMed, 1 May 2021,
<https://pubmed.ncbi.nlm.nih.gov/34212055/>. Accessed 30 July 2025.
- “Mesenchymal stem cell-based therapy for burn wound healing.” PubMed Central,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC8240555/#f1>. Accessed 30 July 2025.
- Rizzello, Giacomo. “Growth Factors and Stem Cells for the Management of Anterior Cruciate Ligament Tears.” PubMed Central, 2012,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3522096/>. Accessed 2025.

“Very nearly... Equal in Status to Caucasians:’ The Mississippi Delta Chinese’s Social and Racial Transformation (1890-1960) By Brendan Wong

Abstract

This paper examines the Chinese population residing in the Mississippi Delta between 1890 and 1960, exploring how they earned a living and interacted with other racial groups, particularly African Americans, in the area. We investigated how the different racial groups interacted related to each other as time progressed. Using oral histories and other records, this paper traces both individual case studies and group trends. The analysis revealed that the Chinese served as middlemen, having good relationships with Blacks in the beginning, and later on with the Whites. Soon after the Chinese came to the Delta, they opened grocery stores in Black neighborhoods, allowing Blacks easy access to goods, earning them good relations with the Black community, as previously, there wasn’t any group willing to help them. However, their relationship soon soured due to how early-20th-century Whites classified the Chinese as colored, impacting their ability to enter better schools and staining their social standing. The Chinese then decided to try to establish better relations with the Whites by completely separating from the Black community to prove to the White community their worth and pseudo-White standing.

In the mid-19th century, people from all over China came to America as a means to earn quick money. They had heard about the railroads and the California Gold Rush in the 1850s and rushed over to make as much money as possible, but came to America with the intention of moving back to China after making enough money to satisfy themselves. Many were contract laborers for the railroads, so after their contracts ended and the railroads were completed, they went back to China. However, while the Chinese laborers worked tirelessly in both hot and cold conditions, they were paid less and treated worse than the White workers²⁶. There were also a select few who were contract laborers for the railroads and then went on to capitalize off of the California gold rush as well.

While this was the case for most sojourning Chinese men, according to one journal written in 1922, there was one Chinese man who moved to Mississippi in search of more work and more money.²⁷ He went by the name Mr. Wong, and, like lots of other young Chinese men at the time, he had come from Canton to California as a contract laborer on the railroads. After he was done in California, he was later shipped out to New Orleans in 1875 to work on a Mississippi riverside-wheeler. After a while, he managed to save up enough capital to start a grocery store in Rosedale, Mississippi. News of his success in the Delta spread, and soon other Chinese men began to move there from both China and California to try and chase financial success.

²⁶ Erika Lee, *America for Americans : a History of Xenophobia in the United States* (Basic Books, 2019), 90.

²⁷ Robert W. O'Brien, "Status of Chinese in the Mississippi Delta," *Social Forces* 19, no. 3 (1941): 386, <https://doi.org/10.2307/2570743>.

After learning of Mr. Wong's success in the Delta, it was said that around 180-200²⁸ Chinese moved to and lived in the Delta by 1900. Like him, a lot of them had previously worked in California on the railroads and then moved out to Mississippi. However, many had shipped straight from Canton because Canton was a port city, making travel much easier than an inland city. Similar to the men from California, the men coming straight from Canton all had the same idea of moving to the U.S. for a short time and ultimately returning to China after making enough money.

After the surge of Chinese Americans died down in the Delta, lots of them had to find work. The Chinese who immigrated straight from Canton mostly found work as sharecroppers, as it was an easy job to acquire. This was because White landowners realized that the Chinese could be "exploited even more effectively" than the Black population, as historian James Loewen argued.²⁹ Eventually, however, the Chinese realized that they wouldn't be able to succeed economically and soon left the sharecropping scene. While the newly immigrated Chinese tried to figure out the sharecropping situation, the Chinese who moved from California were quick to open up grocery stores all around the Delta as their new way of making money.

They quickly opened up grocery stores and became grocers for several reasons that involved the region's economic and social structures, as well as the limited availability of alternative jobs and opportunities at that time. The Chinese initially were only able to find jobs on farms as sharecroppers and field workers. Jobs like bankers, lawyers, doctors, and other higher-paying, higher-class jobs were out of reach for the Chinese. Therefore, they were only able to score jobs in the merchant business or field work. Ultimately, their intense ethic of hard work among the Chinese helped them build up their entrepreneurial pursuits. They did not shy away from working from sunup till sundown, and that is exactly what they did. According to Laura Jue, whose parents first came to America in the 1920s, their store was open till "2-3 am on Saturday nights"³⁰ to accommodate the Black sharecroppers who got off work late and had to go shopping after. Another resident of the Delta said that his family's store would close around midnight and he also noticed a lot of other stores would open at "say 5:30"³¹ in order to "catch the people going out to the fields to work"³² These are just a few examples of how the Chinese attributes and adaptations allowed grocers to economically and socially succeed. Expanding on the social success, Chinese merchants most likely had, at the very least, decent relationships with

²⁸ James W. Loewen, *The Mississippi Chinese : between Black and White* (Harvard University Press, 1971), 25, accessed August 18, 2025, <https://archive.org/details/mississippichine0000loew/page/176/mode/2up?q=symbolic>.

²⁹ Loewen, *The Mississippi*, 30.

³⁰ Bobby Jue and Laura Jue, "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.

³¹ Fay Dong and Juanita Dong, "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster and Jennifer Mitchell, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.

³² Dong and Dong, "Oral History," interview, Delta State.

Black field workers, as they were the only stores willing enough to open early and stay open until very late to accommodate the field workers' long working hours.

These stores were able to help the Chinese climb the social ladder for a few different reasons. For one, the Chinese were indiscriminate about where they opened up their shops, with many Chinese folks opening up in predominantly Black neighborhoods. This allowed them to serve the Black community, which the White merchants largely avoided, as it would tarnish their reputation and usually result in them being kicked to the curb by other Whites. Because of this, the Chinese were able to amass a monopoly on the Black community simply by selling goods to them. This monopoly led lots of Mississippi store owners to earn salaries "often around the five-figure mark"³³ as well as "simultaneously in recent years sending his children through five years of college"³⁴. When considering that, according to 6 different oral histories, the average number of children per family was about 6 children, sending all their children to college becomes even more astounding³⁵.

The Chinese also climbed the social ladder through their frugality to improve their living conditions. While they had no problem spending their money to send their children to school or using it to fight the school boards to send their children to better schools, their homes were often smaller and shabbier than the average merchant of that caliber³⁶. In fact, their houses were often found on the floor above their store or just in a back room. It was common for their stores to be very small, one-room stores, as they were sojourners and didn't plan on needing a large space to operate out of since they would leave the U.S. quickly.

Nonetheless, while they were in the U.S., their small, shabby storefronts offered a very welcoming environment suitable for hangouts for the general neighborhood. They were so

³³ Loewen, *The Mississippi*, 36.

³⁴ Loewen, *The Mississippi*, 36.

³⁵ Kit Gong et al., "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster and Jennifer Mitchell, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Penney Gong, "Oral History of the Chinese in the Mississippi Delta," interview by Georgene Clark, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Edward Joe and Annette Joe, "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster and Jennifer Mitchell, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Bobby Jue and Laura Jue, "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Juliet Kossman, "Oral History of the Chinese in the Mississippi Delta," interview by Jennifer Mitchell, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Audrey Sidney, "Oral History of the Chinese in the Mississippi Delta," interview by Kimberly Lancaster, Delta State, accessed August 22, 2025, <https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>; Dong and Dong, "Oral History," interview, Delta State.

³⁶ Loewen, *The Mississippi*, 109.

welcoming that people would often “meet friends, drink a beer, or leave children while doing errands,”³⁷ as described by historian James W. Loewen. Thus, Blacks and Chinese during the 1900s to 1940s enjoyed relatively friendly relations, with the Blacks comfortable enough to interact with the Chinese, hang out at their stores, and even leave their children there unattended while they ran out to take care of things. Additionally, they were probably very grateful that the Chinese were willing to open their stores so early in the morning and keep them open until very late at night, allowing them to pick up breakfast, shop, and eat dinner, among other things. Furthermore, according to oral histories of the children of Chinese shop owners, many Black customers felt that Chinese merchants “were better to them and kinder to them than a lot of people were”³⁸ and that they “didn’t show that much discrimination”³⁹. This allowed for increasing good relations between the two parties, as the Black workers felt increasing respect coming from the Chinese merchants.

However, while the Blacks had increasingly friendly relations with the Chinese early on, the White folks in town treated the Chinese similarly to how they treated other non-Whites. The Chinese ended up facing similar discrimination as Blacks and other races in the Delta. This discrimination included segregating them into different neighborhoods⁴⁰, denying them basic legal protection from thieves and other businessmen⁴¹, and excluding them from many public schools⁴².

This is because in the late 19th to mid-20th century, according to White social norms, anything that wasn’t White was automatically entered into the colored or non-White category. Because of this, they ended up treating the Chinese without any respect and generally tried to avoid them and their stores. Generally speaking, all the White people in the Delta tended to avoid going to or being seen near Chinese stores because if they were seen there, then they would be considered low status⁴³.

However, over time, both races' perception of the Chinese slowly changed. In the Whites’ eyes, while the Chinese were still excluded from public White schools and were still considered as a low-status group, they were considered as something that wasn’t White but also wasn’t Black, something in the middle. This was a step up for the Chinese, as not being associated with the Blacks as much often came with less discrimination and better relationships. In the eyes of the Blacks, the Chinese were seen as something that wasn’t White but also wasn’t Black. While it distanced the two groups, relations didn’t die down immediately, as the Chinese still provided essential services for the Blacks as they opened shops in Black neighborhoods and acted as middlemen between the White and the Black communities.

³⁷ Loewen, *The Mississippi*, 60.

³⁸ Dong and Dong, "Oral History," interview, Delta State.

³⁹ Jue and Jue, "Oral History," interview, Delta State.

⁴⁰ Loewen, *The Mississippi*, 81.

⁴¹ Loewen, *The Mississippi*, 34.

⁴² O'Brien, "Status of Chinese," 387.

⁴³ Loewen, *The Mississippi*, 50.

The Chinese managed to develop a distinct sense of community through the creation of their own organizations, schools, churches, and other institutions. These efforts were largely aimed at combating the intense racism they faced when first moving to the Delta. The church and its mission allowed the Chinese – at least those Chinese who converted to Christianity – to prove to the White people that they weren't ungodly and immoral, but rather they worshiped the same god and the same things, making them one step closer to the Whites. Additionally, the church provided a space where the Chinese could hold different events like banquets, festivals, weddings, funerals, and community projects⁴⁴. In short, the church provided a space where the Chinese could both prove that they belonged in a higher racial position and hold social gatherings and informal community events.

While the Chinese were able to create a sense of community in the Delta, tensions arose when they attempted to do anything outside of their community. into the existing education system proved especially challenging. The Chinese found difficulty when trying to send their children to White schools. The White establishment, as of 1925, roped Chinese into the colored category and, therefore, only allowed them to attend the Black schools⁴⁵. Chinese families put a lot of emphasis on good education, and because the Black schools were known to be second-rate schools and not the best of the best, they tried to figure out different ways to get better educations for their children. One common way for Chinese children to get a better education than what was provided in the Delta was to either move out of the state entirely or just send the children to live with aunts and uncles in different states⁴⁶. Other parents hired private tutors and homeschooled their children. These methods were the Chinese way of working around the existing segregation and exclusion.

There was one family, however, that didn't agree with their children's education situation and decided to take the county school district to court. The Lum family, and their daughter, named Martha Lum, had been expelled from a White public school in Dublin, Mississippi. Martha, who was born in Mississippi, had previously attended Sunday school before being admitted to Rosedale Consolidated High School, the local White public school. However, during recess, she was notified that she would no longer be allowed to return⁴⁷. Her father, Gong Lum, who held a "considerable standing in the White community"⁴⁸, according to James W. Loewen, went to an additional reputable law firm and hired three lawyers, as well as filing suit against the trustees of the school.

He argued that his daughter was neither a member of the dominant White race nor a member of the colored race either.. She was in her own category, being one hundred percent pure Chinese. Furthermore, he argued that since there were not any schools that were considered 'Chinese only' schools, "separate but equal facilities were not provided for her"⁴⁹. The

⁴⁴ Loewen, *The Mississippi*, 84.

⁴⁵ Loewen, *The Mississippi*, 115.

⁴⁶ Loewen, *The Mississippi*, 68.

⁴⁷ Loewen, *The Mississippi*, 66.

⁴⁸ Loewen, *The Mississippi*, 66.

⁴⁹ Loewen, *The Mississippi*, 67.

Mississippi circuit court for the first judicial district of Bolivar County ended up deciding in her favor, which sparked strife among the school officials, who later appealed to the Mississippi Supreme Court, which reversed the decision, stating that the 1890 Mississippi constitution declared that “separate schools shall be maintained for children of the White and colored races”⁵⁰. Furthermore, the Court stated that the Chinese were not White and therefore must fall under the colored category, which meant that under no circumstances could they attend the all White school. Lum, now joined by another law firm, then challenged the ruling in the United States Supreme Court, which ruled against them by stating, “It has been at all times the policy of the lawmakers of the Mississippi to preserve the White schools for members of the caucasian race alone”⁵¹.

Because of the Supreme Court's ruling, the Lums were forced to either attend a Black school or find a different way of providing education for their children. Ultimately, they ended up moving to Arkansas to provide better education to their two daughters. However, other Chinese started to combat the racist exclusionary schools by making their own schools. Because the Chinese were excluded from White schools, and simultaneously didn't want to attend underfunded Black schools, they banded together and started to fundraise for the creation of Chinese schools.

They were able to secure funds through a few different methods. First, the Chinese Baptist church and its missions set up fundraisers. Second, wholesalers agreed to return one percent of their sales to Chinese workplaces and stores to support this project. And third, once the schools were established, school boards agreed to utilize some of the county's tax money to hire White teachers⁵².

The Chinese desire not to integrate into the Black schools stemmed from two reasons. First, it was well known that the education that was available at the Black schools was far inferior to the White schools, as White schools got “283,161 dollars for 6216 White students and 38765 dollars for 35708 Black students”⁵³. Secondly, the Chinese wanted to start disassociating with Black people and Black culture, as they increasingly wanted to join the Whites in the upper echelons of society.

The Chinese were further pushed to try to assimilate into White society because of the xenophobia of the early 20th-century U.S. The anti-immigrant and especially anti-Asian sentiment was reflected by the immigration restrictions put into effect during the late 1800s to mid-1900s to try to sift through and only let in White or Western Europeans. This xenophobia, in conjunction with eugenics, led the Chinese to attempt to assimilate with the Whites in order to gain more social standing, but also to skirt around the immigration restrictions and general discrimination. Additionally, around the 1920s, the United States experienced the resurgence of the KKK. This time, the newly resuscitated KKK opened its horizons and not only targeted

⁵⁰ Loewen, *The Mississippi*, 67.

⁵¹ Loewen, *The Mississippi*, 67.

⁵² O'Brien, "Status of Chinese," 388.

⁵³ Loewen, *The Mississippi*, 65.

African Americans but also other ethnic and immigrant groups⁵⁴. This gave the Chinese another reason to try and gain social standing with the Whites as soon as possible, as they did not want to fall victim to everything that was happening around them and to their peers.

The Chinese had created a deliberate and multifaceted strategy in order to make themselves seem more White. They wanted to move away from being classified as colored and similar to Black status and move closer to the status of Whites. Their desire not only led to their separation from the Black schools, but it also led to the exclusion of all Black and Black-Chinese in their community. When the Chinese men moved from Canton to the Delta, lots of them couldn't bring their families, which included wives and children, with them. Because of this, lots of Chinese men married Black women and subsequently had children with them. But by the mid-1920s, as the broader Chinese community wanted to move away from the Black classification, many had to choose between their Black or part Black families or the Chinese community. If they chose their families, the Chinese community would ostracize them, but if they chose the community, they would need to ostracize their families⁵⁵. It was a difficult point that turned the tide of Chinese-Black relations.

By the 1940s, another way the Chinese tried to make themselves seem more White or strengthen their relationship with the Whites, and specifically the White upper class, was to fight in World War Two. Their positions as allies in the war and, for some, maybe even heroes, boosted their popularity back home with White people, helping them to again seem better than the Black community. Additionally, since lots of White men fought in the war as well and had to travel overseas when they came back, some Chinese folks felt that since they had travelled and seen things, their attitudes towards the Chinese and other races had changed⁵⁶.

After the Chinese successfully removed themselves from the Black classification, their lives and social interactions were significantly transformed. By the 1930s, they were seen as an intermediate group, not quite White, but far from Black. They still weren't allowed to attend White public schools, but they got their own separate school buildings. By the 1950s, however, they were "very nearly, and in some respects entirely, equal in status to caucasians"⁵⁷. While the Chinese have the same or very near status as the White people now, it didn't come with sacrifice. In order to attain their new and improved status, the Chinese had to sacrifice their customs and culture.

One such thing that was lost over time was language. As would be expected, the Chinese who immigrated from China didn't speak English, which forced them to converse in different regional dialects with their children, neighbors, and friends. However, as they started to try and classify themselves as White, they stopped trying to teach their children Chinese as they wanted them to seem more White. Fay Dong, a second-generation Chinese American who grew up in the Delta, noted this in her oral history from May 1, 2000, that she could "get by" speaking Chinese but never learned to read or write. Furthermore, she recounts that her children learned Chinese

⁵⁴ Lee, *America for Americans*, 15.

⁵⁵ Loewen, *The Mississippi*, 143.

⁵⁶ Dong and Dong, "Oral History," interview, Delta State.

⁵⁷ Loewen, *The Mississippi*, 2.

from their grandmother, who refused to speak English, but, similar to her, weren't taught how to read or write.

Another thing that was lost to the Chinese assimilation into White culture was parental authority over their children's social lives and structures⁵⁸. As time passed, more and more Chinese youth wanted to embrace the American form of romantic love through dating and marriage, as well as expanding their social circles to more than just other Chinese. While they didn't obey their parents' rules and ideals on marriage, most Chinese youth did pursue higher education, with a vast majority of Delta Chinese ending up going to college and choosing successful fields like engineering and pharmacy, which allowed them to get higher-paying starting jobs⁵⁹.

Although they were able to achieve such high academic and professional success, most, if not all, of the Chinese had to move out of the Delta. This was because the Delta was simply too small for the number of Chinese with degrees in high-paying professions. Therefore, they needed to move out and across the U.S. in order to find the right place and job. Unfortunately, because lots of children moved from the Delta to other states, the once family-owned Chinese grocery stores had to close down because their parents couldn't keep up⁶⁰ with the demand as they grew older.

However, the few that were able to survive soon closed due to the civil rights era of the 1960s. As the Chinese gained acceptance into the White community, their relationship with the Black community became increasingly strained. The Blacks saw the Chinese gaining their improved status and saw it at their own expense. The way the Blacks saw it, the Chinese were aligning themselves more and more with the White social group and leaving them to be the only ones in an oppressed state⁶¹. The Blacks didn't take this lightly, and during the civil rights era, they showed their distaste with the Chinese⁶².

Because they saw the Chinese as White, Black advocates for emancipation often targeted the Chinese stores for a handful of reasons. Chinese stores were mostly found in Black neighborhoods, making them easy targets for Blacks to destroy, and to do that, they often used "bricks and firebombs"⁶³ to break windows and destroy the stores. Another reason for the targeting of Chinese stores was that although the Chinese were considered White, they couldn't mobilize the police as quickly or effectively as White merchants⁶⁴. Furthermore, these attacks served, in Loewen's perspective, as a "symbolic attack upon the White community, which is the ultimate target,"⁶⁵ in which the Chinese acted as surrogates for Whites.

The Chinese population in the Delta began to decline slowly throughout the 1950s and 60s, primarily due to a decrease in the number of grocery stores. The grocery stores started

⁵⁸ Loewen, *The Mississippi*, 158.

⁵⁹ Loewen, *The Mississippi*, 179.

⁶⁰ Jue and Jue, "Oral History," interview, Delta State.

⁶¹ Loewen, *The Mississippi*, 138.

⁶² Gong, "Oral History," interview, Delta State.

⁶³ Loewen, *The Mississippi*, 175.

⁶⁴ Loewen, *The Mississippi*, 176.

⁶⁵ Loewen, *The Mississippi*, 176.

decreasing for a few notable reasons⁶⁶. One main reason was that in the 1940s and 50s, the Delta saw a boost in industrialization, mainly in the farming industry. Because the Delta economy heavily relied on cotton farming, the fact that lots of farms were industrializing meant less need for actual workers, and since most of the workers were Black, that meant that the Black population in the Delta took a heavy hit. This was especially brutal for the Chinese since their primary customers were all Black farm workers. Furthermore, since education was valued so highly in their families, Chinese children were starting to expand and go out of state in order to pursue higher education. Part of the reason was that they didn't want to be like their parents and get stuck in the repetitive life of being a Delta grocer. Another reason that most Delta children don't want to come back is that they believe that there's nothing left where they grew up⁶⁷, and therefore, no room for growth.

While the Chinese in the Delta have mostly died out by now, various groups have kept alive an interest in the history of the place as well as the people. For example, there is a Facebook group called Mississippi Delta Chinese R Us⁶⁸ that enjoys posting about new discoveries and revelations found out about specific people and groups that lived in the Delta at the time.

The Chinese in the 20th-century Delta exemplified the complexity between two "inferior" groups in the area, and how they both utilized each other to strive for betterment. However, they proved that mutual inferiority does not always equate to perpetual solidarity. The breakdown of solidarity only occurred because the Chinese sought a better social standing, which they believed required leaving the Black community to fend for itself. In order to maintain a stance of solidarity, both races would have required good social standing and incentives for keeping good relations. The downfall of solidarity between the groups came when the Chinese felt they hadn't achieved their dream of making a better life in the U.S.. This realization caused them to betray the Black community and try to side with the dominant White population, which ultimately led to ruining the relations between the two "inferior" groups.

⁶⁶ Loewen, *The Mississippi*, 172.

⁶⁷ Jue and Jue, "Oral History," interview, Delta State.

⁶⁸ <https://www.facebook.com/groups/100371283356445/>

Works Cited

- Anderson, Zora, Louise Witt, Jeff Bogue, Ho Shee, Louise Simpson Hamilton, and E. B Reynolds. "An Interview with the Teachers of the Chinese School." By Daisy Greene. Mississippi Department of Archives and History. Accessed August 22, 2025.
<https://da.mdah.ms.gov/series/audio-collections/transcripts/detail/1051776>.
- The Delta-Democrat Times (Greenville Mississippi). "Negro Held in Store Burglary." April 17, 1937. Accessed August 19, 2025.
<https://www.newspapers.com/article/the-Delta-democrat-times/174981702/>.
- The Delta-Democrat Times (Greenville Mississippi). "Officers Watch for Highwayman." September 22, 1932. Accessed August 19, 2025.
<https://www.newspapers.com/article/the-weekly-democrat-times/174981672/>.
- The Delta-Democrat Times (Greenville Mississippi). "Youth Bound over for Chinese Grocery Robbery." August 18, 1939. Accessed August 19, 2025.
<https://www.newspapers.com/article/the-Delta-democrat-times/174981638/>.
- Dollard, John. *Caste and Class in a Southern Town*. 3rd ed. Doubleday Anchor Books, 1949. Accessed August 19, 2025.
<https://www.scribd.com/document/458115121/Caste-And-Class-In-A-Southern-Town-By-John-Dollard-pdf>.
- Dong, Fay, and Juanita Dong. "Oral History of the Chinese in the Mississippi Delta." Interview by Kimberly Lancaster and Jennifer Mitchell. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Edwards, Paul K. *The Southern Urban Negro as a Consumer*. Negro Universities Press, 1969. Accessed August 19, 2025.
<https://www.loc.gov/resource/gdc.71084098/?st=slideshow#slide-6>.
- Gerstle, Gary. *American Crucible : Race and Nation in the Twentieth Century*. Updated ed. Princeton University Press, 2017.
- Gong, Kit, Bobbie Gore, Joy Gore, Amy Gore, and Billie Gore. "Oral History of the Chinese in the Mississippi Delta." Interview by Kimberly Lancaster and Jennifer Mitchell. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Gong, Penney. "Oral History of the Chinese in the Mississippi Delta." Interview by Georgene Clark. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Holley, William C., Ellen Winston, and Thomas Jackson Woofert. *The Plantation South, 1934-1937, Research Monograph XXII*. Washington DC: U.S. Government Printing Office, 1939. Accessed August 19, 2025.
<https://fraser.stlouisfed.org/title/plantation-south-1934-1937-6707?page=5>.

- Joe, Edward, and Annette Joe. "Oral History of the Chinese in the Mississippi Delta." Interview by Kimberly Lancaster and Jennifer Mitchell. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Jue, Bobby, and Laura Jue. "Oral History of the Chinese in the Mississippi Delta." Interview by Kimberly Lancaster. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Kossman, Juliet. "Oral History of the Chinese in the Mississippi Delta." Interview by Jennifer Mitchell. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- Latourette, Kenneth Scott. *The Chinese : Their History and Culture*. Macmillan Company, 2005.
- Lee, Erika. *America for Americans : A History of Xenophobia in the United States*. Basic Books, 2019.
- Loewen, James W. *The Mississippi Chinese : between Black and White*. Harvard University Press, 1971. Accessed August 18, 2025.
<https://archive.org/details/mississippichine0000loew/page/176/mode/2up?q=symbolic>.
- O'Brien, Robert W. "Status of Chinese in the Mississippi Delta." *Social Forces* 19, no. 3 (1941): 386-90. <https://doi.org/10.2307/2570743>.
- Percy, William Alexander. *Lanterns on the Levee; Recollections of a Planter's Son*. with an Introd. by Walker Percy. New York: Knopf, 1941.
- Rose, Harold M. "The All-Negro Town: Its Evolution and Function." *Geographical Review* 55, no. 3 (1965): 362-81. <https://doi.org/10.2307/213134>.
- Sidney, Audrey. "Oral History of the Chinese in the Mississippi Delta." Interview by Kimberly Lancaster. Delta State. Accessed August 22, 2025.
<https://www.Deltastate.edu/library/departments/archives-museum/guides-to-the-collection/oral-history-collections/Chinese-oral-histories/>.
- The Weekly-Democrat Times (Greenville Mississippi). "Holdup of Chinese Grocery Is Foiled; Negro Hunted Here." February 24, 1938. Accessed August 18, 2025.
<https://www.newspapers.com/article/the-weekly-democrat-times-Chinese-Black/174292897/>.

Gender Inclusion in STEM: Feminist Perspectives on VEX Robotics By Anya Li

The K-12 educational experience is increasingly dominated by participation in competitive extracurricular activities, from physical sports to debate teams to robotics. Among the key issues facing student organizations is the role of diversity and equity in competitive spaces. This essay uses feminist theories of organizational communication to examine the gender inclusivity efforts of VEX Robotics, an international high school competition. VEX has implemented a wide range of inclusion efforts, yet still faces routine criticism from female participants who argue that VEX competitions are rife with toxic masculinity. This essay first considers the application of foundational feminist theories to the extracurricular competition space, then evaluates the choices made by VEX to achieve meaningful gender inclusion and equity. Specifically, I use a combination of ethnographic and autoethnographic rhetoric to evince the experience of women in VEX as a means of illuminating the continued struggles of female participants. While many institutional changes would suggest VEX to be above the curve when it comes to inclusivity, resistance is baked into the institutional framework of competitive activities, which yields continued challenges for women.

More students than ever are involved in multiple competitive outlets, from physical sports to debate teams to robotics. Among the key issues facing student organizations is the role of diversity and equity in competitive spaces (Buzard, Gee, Stoddard). As these competitions have ostensibly neutral criteria for success, organizers face difficult decisions in how to promote diversity and level the playing field without undermining the spirit of fairness. Within that space, many organizations execute diversity and inclusion initiatives that have mixed results (Silver).

Potential contradictions between the tenets of inclusion and competitive equity mirror the tensions faced by feminist theorists in organizational spaces. Many workplaces, institutions, and organizations may desire increased representation and empowerment of women, but will limit programs for advancement to avoid criticism of bias or out of obligation to profit maximization. Many of the same barriers faced by women in the workplace are replicated in coed extracurricular competitions. Despite efforts to prioritize female participation, competitions are limited in the mechanisms for empowerment and must be cautious to avoid perceptions of bias that may threaten fairness. For example, robotics competitions could not award additional points, time, or attempts to teams simply for including more female participants.

This essay uses feminist theories of organizational communication to examine the gender inclusivity efforts of VEX Robotics, an international high school competition. VEX Robotics is a program for students from elementary through university to design robots and compete in competitions managed by the Robotics Education & Competition Foundation (About Us). VEX Competitions are the largest robotics program with around 1.1 million students participating worldwide. In VEX V5, a branch of VEX Robotics targeting middle and high school, there are more than 20,000 teams from 50 countries. It consists of over 1,700 competitions each year (RobotEvents).

VEX Robotics has taken visible steps toward gender inclusion. While the organization has promoted female participation and showcased women's leadership, existing tensions between genders persist in competition activities. At the executive level, the REC Foundation's Board of Directors includes Dan Mantz, Dr. Cori Lathan, Mike Harris, Mary Lou Ewald, and Dr. Shelly Gruenig, two men and three women (About Us). This represents a positive balance of gender representation and signals some inclusion of women's ideas. However, VEX Robotics was originally founded by two men, Tony Norman and Bob Mimlitch, and the culture of robotics has long been male-dominated. Over the years, there has been a consistent pattern of around 65% male participants compared to around 31% females in VEX (Sullivan & Bers). The question remains as to why the system continues to produce such a wide gender gap.

While existing scholarship on VEX has examined participation rates, gender attitudes, and biases, I will contribute from a feminist standpoint theory, as well as examining the perspectives of participants themselves. Specifically, I consider the gender inclusive program Girl Powered and patchwork efforts within VEX to improve the number of female participants. To address this, I will analyze VEX's existing inclusivity initiatives using the FST and autoethnographic perspective as a long-time participant. I evince my claims by examining official VEX materials as well as ethnographic and autoethnographic discourse. Empowering the voices of actual participants will illuminate the downstream effects of VEX initiatives, highlighting the persisting tension between inclusivity campaigns and structural barriers that limit full equality.

Feminine Perspectives in Communication, Rhetoric, and Organizational Culture

Communication theory has made great strides in our understanding of the role of feminism as an equipment for living (Buzzanell). Rhetorical scholars have interrogated the nature of gender in public address, leadership roles, and great historical moments (McAlister). Organizational communication scholars have transplanted theoretical work to the everyday contexts of labor, interpersonal organizations, and other human activity arrangements (Muhamad). These insights analyze how gender is structured, at both theoretical and practical levels, in institutions as large as the UN and as local as VEX Robotics. I will briefly note a few significant progress points in the intersection of communication and feminist studies to underline the theoretical assumptions at work in my analysis of VEX.

Rhetoric feminist theory provides an understanding of how women navigate male-dominated fields by working in new forms of communication style. Karlyn Kohrs Campbell has shown the tension between traditional roles of women/traditional femininity and the masculine act of public address by illustrating how women must fight against the contextual resistance of audience expectations for demurity and passivity. Campbell argues that female orators "[violate] the reality structure" of permissibility by the very act of speaking in male-dominated arenas (Campbell). Therefore, Campbell argues, women's rhetoric is significant both because of its source and because it must overcome internalized psychic barriers that modulate the texture of women's public address. Women's rhetoric is thus not only the speech of

women, but a method of speaking that begins with a willingness to violate precise cultural taboos. Dow and Tonn take Campbell's work even further by demonstrating how women's rhetoric is not limited to women's issues, but instead reflects a set of practices such as reducing the public to the private or actively traversing the speaker-audience power gap by inverting the authority of the speaker (Dow and Tonn). As foundational works in the study of feminist communication, the theoretical contributions of Campbell, Dow, and Tonn are still relevant in male-dominated fields like VEX because the conditions of exclusion against female participation may differ in magnitude but echo historic forms.

Feminist theory in organizational communication may prove more directly applicable to the working conditions of VEX robotics. The first subset of organizational theory of relevance to my analysis is often classified as "the liberal feminist approach" (Ashcraft 128). This theory states that gender is predetermined and prejudice starts with individuals and leaks into organizations (Benschop). Elsesser and Lever's research shows, for example, that women managers are not fundamentally different from male managers in efficiency or ability, but are nonetheless treated as less competent (Elsesser and Lever). Women in workplace environments often face double standards where similar behaviors are interpreted differently depending on the gender (Eagly & Karau). These double standards cause barriers to advancement in hierarchies since women are held back due to biased ignorance (Tremmel & Wahl). Attempting to adhere to masculine standards in the workforce often causes direct backlash as well when women fail to act "feminine" in the course of duty. Such double binds enforce stereotypes of femininity on women who take on the traditionally masculine roles. (Koenig & Eagly). Unfortunately, in practice, the best option for an individual woman seeking business success is to conform to masculine norms rather than challenge them (Weiner & Burton). Others will take a step better and run workshops or advancement seminars to empower other women to achieve masculine norms, but rarely is there pressure for men to adapt feminine ways of knowing or doing. Liberal feminist organizational theory shows how gender bias is built into environments like VEX Robotics and stems from individual discrimination.

Cultural feminism reads the position of womanhood as a deviation from the cultural hegemonic core of heteromascularity (Syed & Murray). As gender is castigated and defined through difference, cultural feminist critique enables us to see the individual experience as part and parcel of a systematic effect, nonetheless an effect that can be similarly mediated and intervened from the individual perspective (Rodriguez & Guenther). Masculine styles of communication are norms in organizations, while feminine styles are undervalued and harmful for women (Tannen). For example, Tannen argues that communication practices such as conflict, leadership, self-promotion, and strong versus weak speech are all deemed relevant to work and are seen as masculine norms. Ashcraft further claims that programs train women to be more assertive, assuming male norms are ideal, and try to make women more like men instead of being based on their own communication methods. These programs erase the possibility of other communication styles and undermine the possibility of broadening effective leadership with feminine communication (Syed & Murray). Cultural feminism applied to VEX will reveal a

pattern of well-intentioned but potentially calamitous interventions in the form of programs that aim to instill masculine strengths in female participants.

Radical poststructuralist feminism argues that organizations are designed in male-centered ways and that gender is the apex difference, so all systems are inexorably founded on gendered binaries (Ferguson). Ferguson argues that even women-led organizations often copy patriarchy because gender hierarchy is a set of practices, not the inevitable result of gendered persons. Furthermore, Ferguson shows how the success of individual women should not be confused with structural advancements in the conditions of possibility for equality. For example, Ashcraft and Pacanowsky warned that a workplace might celebrate women's communication style, but still measure success through masculine criteria, ultimately creating more granularity in the acceptable range of feminine behaviors, but only insofar as those behaviors are residually masculine. Acker showed that gender is often built into job roles and runs on masculine norms regardless of who holds the position, resulting in situations where female bosses or CEOs fail to implement top-down institutional changes because of contextual constraints (e.g. shareholder demands, fiduciary duty, market forces). This reading of gender would help us understand why VEX may fail to achieve equal gender participation despite robust inclusion efforts and may point to systematic limitations on intervention.

Feminist Standpoint Theory (FST) in organizational communication pushes these critiques further by emphasizing the importance of evaluating structures from the position of the dispossessed (Gurung). FST has been used to analyze how race, gender, and status shape one's experience (Maake). This theory claims that differences stem from the nuances of social location (the localized cultural, political, and economic spaces that constitute lived reality), instead of universalized masculine or feminine cultural cues (Crasnow). Allen illustrates how Black feminist standpoints challenge existing norms surrounding professional socialization, including in academic organizations. For Allen, a key area for intervention is the social gaps between institutional rails where individuals are empowered to make interpersonal decisions. Accordingly, those interpersonal acts may be implicitly racialized or gendered even when the actor is well-intentioned. Similarly, Parker draws from interviews with 15 African American women to illustrate the unique difficulties faced by intersectional identity that may be overlooked in either strictly gendered or strictly racialized analysis (Parker). These examples show the importance of the broader FST framework that treats the perspectives of historically marginalized groups as sources for reshaping organizational practices. They provide alternatives to dominant white masculine norms, such as collaborative decision making and diverse leadership practices. FST colors the solution by offering a guideline for intersectional approaches to equity that minimizes the reinforcement of existing power dynamics across other identity categories. Rather than assimilating women to fit masculine norms, FST changes the way we define difference and leadership (Gurung). In VEX, FST could help reestablish team cultures that reflect the perspective of the girls.

Current research on gender and STEM reveals that stereotypes, confidence, and skill perception are widely different between men and women, further shaping the gap between

participation in STEM. Existing research has confirmed that, for many, STEM disciplines are considered masculine and that high intellectual abilities that seem necessary in STEM are also more common in men (Carli et al.; Miller et al.; Leslie et al.; Roberts & Light). Men have reported feeling more confident in STEM even when the results from women greatly surpass men (Bian et al.; Roberts & Light). The results seen in the general population can also be reflected in robotics. In a study focused on the World Robot Olympiad, all-girl teams had advantages in communication and collaboration but lacked in their robot-building skills (Chiang et al.). These embedded beliefs can cause women's lowered confidence in robotics and engineering (Roberts & Light). Sullivan and Bers's research on genders' attitudes in VEX Robotics is consistent with previous STEM-wide studies, finding that boys feel more confident in their general technical ability, whereas girls feel more confident in areas such as writing, scouting, and engineering notebooks. Internal VEX surveying reveals that female participants have the same level of interest and enthusiasm for robotics as their male peers, yet are still often relegated to passive roles in actual team compositions. A VEX Robotics competition evaluation conducted by Cher Hendricks shows that girls' interest in STEM is not the issue. Specifically, 96.2% of girls stated that VEX participation made them want to learn more about robotics, compared to 91.8% of boys (Hendricks et al.). Similarly, 78.5% of girls became more interested in additional STEM classes versus 74.9% of boys (Hendricks et al.). In both scenarios, girls have shown more interest in STEM than boys.

Feminizing VEX Robotics

Despite demonstrable interest, women in VEX are not always given the same opportunities to express that interest within the team dynamic. I evidence this claim with previous research, quotes from VEX online discussion boards, and personal experience as a woman in VEX.

Women in VEX are often at a disadvantage due to the perception that women are less capable at engineering than men, resulting in disempowerment that discourages active participation. According to quotes from team mentors, girls are often expected to take a more passive role on the team, whereas males take the lead even if they do not necessarily know more about the task (Sullivan & Bers). Male students report significantly higher confidence in both technical ability and putting things together, which means self-selected team leaders will be predominantly male. Since women show higher confidence in collaborating with team members, the perception of male superiority in leadership roles emerges from an overly narrow understanding of team contribution. This conflagration of role capacity harms the entire team, often disempowering women from the early stages of development. Mentors often feel the need to give girls more of a push to maintain team dynamics, resulting in the perception that women require additional coaching to meet minimal standards. Friends of mine on mixed gender teams have reflected these patterns, describing how the boys frequently insist on building or fixing the robot themselves, leaving little room for female contribution, which in turn causes coaches to question the female members' commitment and efforts. Despite coaching often zeroing in on the

efforts of excluded participants, VEX mentorship recognizes that exclusion is often maintained socially. One mentor states that “guys will exclude girls if not shown the value of her as a teammate” (Sullivan & Bers).

Despite recognition of systemic effects of discrimination, inclusion efforts often focus on individual participant self-perception and confidence rather than the structural causes of inequality. Even Sullivan and Bers recommend that mentors should focus on enhancing female students' confidence to get on the same level as the males. Similarly, in my own experience at girls' events such as the Girl Powered summer workshop hosted by Google, the emphasis was on making us comfortable with building and coding by giving step-by-step manuals and pre-written base code. This left participants more passive in their interaction than in a typical VEX experience, as female students were left to follow instructions instead of experimenting with building or programming. These initiatives, though well-intentioned, reinforce cultural feminist critiques by demonstrating a focus on making women fit into masculine archetypes with little to no consideration given to changing men's behavior to better fit feminine expectations.

Across all levels of competition in VEX, gender disparity remains significant, showing that teams still tend to reflect male-dominated norms. The issue of stereotypes causes “strategic shrinking”, where teams hesitate to include girls, whether that's in their team or during alliance selection at competitions (Sullivan & Bers). Without structural incentives or mandates, maintaining equity might be difficult. Data from a VEX Robotics Competition Evaluation highlights that this bias has already been shown in team composition. In one study, only 1.5% of teams reported having an all-girls team, whereas 16.5% reported having an all-boys team (Hendricks et al.). Similarly, around 67% of the Team Leaders/coaches were male. From the feminist standpoint, these statistics reveal how individual prejudice has affected the structural imbalance. Although female participants show an equal interest in robotics, they are less likely to be in leadership positions because teams unconsciously replicate existing gender norms.

Within many mixed-gender teams or VEX competitions, players tend to conform to stereotypical team roles: boys dominate the coding, driving, and building, while girls are assigned to notebook, scouting, or social media outreach. While this is not universal, it is an expectation that has affected many girls' experiences in VEX since they feel mistreated on their team or at competitions with other boys. In my personal experience on an all-girls team, I have experienced a lot of prejudice from other teams during competitions. In the 2024-2025 season, my team was unique in trying a different drivetrain from the usual robots. During one competition, we overheard another team mock our choice by saying, “What if we just did something stupid like X drive” while glancing at us. Internal reports from a female competitor on that team claim she was often excluded from the building and decision-making, which adds another gendered layer to our interaction. Additionally, we had several occasions where teams openly questioned our build strategies, such as why we bent a C-channel a certain way for our lady brown arm lift system. These interactions were often dismissive, rather than inquisitive. Comments made us feel like our team was incompetent in building, as if we did not know what

we were doing, while other teams seemed to inquire over each other's builds for strategic purposes or intel gathering.

In an online VEX forum started by Ssang, she states, "They talk about a lot of stuff that I don't really follow (cars, Minecraft, D&D), not that girls can't do those things, but I don't, and I don't really want to either. Also, it's annoying when people come in for the first time and automatically assume the boys carry the team, and I just do the notebook". There were mentions of the girls on the team not being able to connect with the boys due to the hobby difference, also an annoyance when it came to people assuming the boys carry the team, and the girl just does the notebook. This example is representative since it aligns with previous findings that boys take on the leadership role and girls feel less confident in the build, so they turn to roles such as notebooker or scout. Other people have also come out to agree with the original poster, stating that boys are sexist towards them, and even the coach gives them the easiest parts of the robot to work on. These anecdotes show that girls want to contribute to the team but are often positioned in more "feminine" roles.

One well-intentioned initiative is Girl Powered, a RECF-supported campaign designed to increase female interest in robotics. Their goal is to "redefine the face of STEM" and provide tools and foster welcoming environments to help girls gain confidence in robotics (Girl Powered Redefining the Face of STEM). Girl Powered creates educational videos, blog posts, and hosts workshops to highlight the benefits of fostering inclusion. Additionally, they encourage teams to submit their own posts of female participation to get more personal stories (Girl Powered Redefining the Face of STEM). At competitions such as the World Championship, Girl Powered handed out key chains and flyers to teams to continue supporting their cause. One of Girl Powered's main events is their summer two-day workshop, where teams of students will get to design and build a robot in one day, then compete with other teams in the given year's challenge (Girl Powered Robotics Workshop 2025). These participants consist of students who have no prior experience in VEX and have the opportunity to hear from those with competitive robotics experience and STEM industry professionals (Girl Powered Robotics Workshop 2025). Girl Powered is revealed to have both strengths and limitations. They provide valuable visibility to help girls access STEM opportunities; however, cultural feminist critiques would suggest that the initiative ultimately reinforces masculine norms by framing inclusion as teaching girls to catch up and build their own confidence.

At the 2025 VEX World Championship, Girl Powered hosted a lunch presentation to celebrate women in STEM. While this event was well-intentioned, it had little effect since it was not advertised well and scheduled at an inconvenient time. My own all-girls team received a personal invitation from the organizers, but many other teams were oblivious to the event. Even for us, attending was impractical since lunch time is prime time for teams to repair or tune the robot; therefore, we had to decline. A more impactful approach would be dedicating time during the official competition schedule to host such presentations, such as during the opening or closing ceremony. This way, most teams would be present and more focused on the presentation, and also signals that inclusion is central and part of the competition.

However, this initiative still often reinforces existing frameworks and emphasizes a group's confidence-building instead of changing the structure of participation. In a Girl Powered presentation given by team 7700R, they believe that the Girl Powered movement means “building girls’ talents and allowing them to show their passions” (Chang). While this initiative has some radical perspectives embedded, there are still beliefs to make women more like men. They suggest that girls need to build their talents to be more confident in VEX. Girl Powered raises awareness of gender disparities and takes meaningful steps toward equity, but it would need to shift from confidence-building to structural redesign in order to fully embody a radical feminist approach. At the same time, Girl Powered has reinforced the double binds that feminist theories critique. They are working to boost confidence, signalling that girls are the problem rather than attacking organizational norms that privilege masculine behaviors. By assuming success in VEX depends on adopting these masculine traits removes the value of collaborative skills important in feminine communication. The program celebrates women’s participation without questioning the underlying framework, reproducing patriarchal thinking even within a women-led framework. In this sense, Girl Powered is benefiting in visibility but falls short in restructuring the underlying norms in robotics.

Conclusion

VEX Robotics has done more than many rival organizations when it comes to the empowerment of women and other efforts towards representational equity. By advancing women to high-ranking positions within the organization, directly supporting female competitors through GirlPowered, and overrepresenting women in marketing materials, VEX has made considerable progress for women in K-12 robotics. Despite this progress, the nature of patriarchy as a social cudgel necessarily yields exclusionary practices that continue to leave women sidelined in their own teams. These effects often result from the dual inputs of stereotyped assumptions and competitive drive. Ironically, such assumptions often lead to underutilizing female teammates, further contributing to the expectation that women are less capable of success in robotics than men. VEX should continue to platform female keynote speakers, celebrate female participants, and fund the development of female teams.

Creating institutional change in competitive activities necessitates balancing fairness and equity. Awarding participation trophies for women does not address the problem of integration nor the problem of peer respect. Improving the scores of teams that include women would incentivize mixed-gender teams, but failure to address the underlying causes of discrimination would most likely lead to all-male teams adding women without integrating them into the team. Nonetheless, there are many knobs for competitive evaluation that can be tweaked to incentivize feminine skills. Sportsmanship awards could be offered that are based on peer evaluation or essay writing contests. These awards would be independent of team scoring but would still be sought after by competitors using VEX as a resume-building activity. Alternative scoring systems could weigh style, design, or decoration as part of the overall robot evaluation. Importantly, sportsmanship or aesthetics would have to contribute in some way to the regional

qualification system for changes to meaningfully influence student behavior. In the interim, individual awards could be provided for *esprit de corps* or congeniality. New awards could be created to reward the social aspect of VEX that reward team social media presence, community involvement, or contest-related volunteering – all wells of emotional labor, disproportionately undertaken by women, and routinely undervalued.

Works Cited

- Acker, Joan. "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender and Society*, vol. 4, no. 2, 1990, pp. 139–158. JSTOR, <http://www.jstor.org/stable/189609>.
- Alemdar, Meltem, et al. "VEX Robotics Competition Evaluation." RECF, 3 June 2011, recf.org/documents/2019/08/study-vex-robotics-competition-evaluation.pdf.
- Allen, Brenda J. "Feminist standpoint theory: A black woman's (re)view of organizational socialization." *Communication Studies* 47, 1996: 257-271.
- Ashcraft, Karen Lee. "Feminist Theory." *The SAGE Handbook of Organizational Communication: Advances in Theory, Research, and Methods* eds Putnam, L. L. & Mumby, D. K. 2013, pp. 127–46.
- Ashcraft, Karen Lee, and Michael E. Pacanowsky. "'A Woman's Worst Enemy': Reflections on a Narrative of Organizational Life and Female Identity." *Journal of Applied Communication Research*, vol. 24, no. 3, 1996, pp. 217–239.
- Benschop, Yvonne, and Mieke Verloo. *Feminist Organization Theories: Islands of Treasure*. Routledge, 2016.
- Bian, Lin et al. "Gender stereotypes about intellectual ability emerge early and influence children's interests." *Science (New York, N.Y.)* vol. 355, 6323, 2017: 389-391. doi:10.1126/science.aah6524
- Bowers, Shalla. "The Future Is Female: Springfield Robotics Beats Gender Statistics." *KHQ TODAY*, 3 Mar. 2023, khqtoday.com/feature/2023/03/03/the-future-is-female-springfield-robotics-beats-gender-statistics/?utm_source.
- Buzard, Kristy, et al. *Who You Gonna Call?: Gender Inequality in External Demands for Parental Involvement*. IZA - Institute of Labor Economics, 2025. JSTOR, <http://www.jstor.org/stable/resrep70870>.
- Buzzanell, Patrice M. "Gaining a Voice: Feminist Organizational Communication Theorizing." *Management Communication Quarterly*, vol. 7, no. 4, 1994, pp. 339–383. SAGE Publications, doi:10.1177/0893318994007004001.
- Campbell, Karlyn Kohrs. "The Rhetoric of Women's Liberation: An Oxymoron." *Communication Studies*, vol. 50, no. 2, 1999, pp. 125–137. Taylor & Francis Online, <https://doi.org/10.1080/10510979909388479>.
- Carli, Linda L., Laila Alawa, YoonAh Lee, Bei Zhao, and Elaine Kim. "Stereotypes about Gender and Science: Women ≠ Scientists." *Psychology of Women Quarterly*, vol. 40, no. 2, 2016, pp. 244–260. SAGE Publications, <https://doi.org/10.1177/0361684315622645>.
- Chang, Samantha. *Girl Powered: The Power in Girls*. REC Foundation, 2024. REC Foundation, https://challenges.robotevents.com/uploads/0014436_original.pdf.
- Chiang, Feng-Kuang, Zhonghua Tang, Dan Zhu, et al. "Gender disparity in STEM education: a survey research on girl participants in World Robot Olympiad." *International journal of technology and design education*, 1-18. 6 May. 2023, doi:10.1007/s10798-023-09830-0

- Crasnow, Sharon. "Feminist Standpoint Theory." *Philosophy of Social Science: A New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi, Oxford University Press, 2014, pp. 145–161.
- Dow, Bonnie J., and Mari Boor Tonn. "'Feminine Style' and Political Judgment in the Rhetoric of Ann Richards." *Quarterly Journal of Speech*, vol. 79, no. 3, 1993, pp. 286–302. Taylor & Francis Online, <https://doi.org/10.1080/00335639309384036>.
- Eagly, Alice H., and Steven J. Karau. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review*, vol. 109, no. 3, 2002, pp. 573–598. American Psychological Association, <https://doi.org/10.1037/0033-295X.109.3.573>.
- Elsesser, Kim M., and Janet Lever. "Does Gender Bias against Female Leaders Persist? Quantitative and Qualitative Data from a Large-Scale Survey." *Human Relations*, vol. 64, no. 12, 2011, pp. 1555–1578.
- Ferguson, Kathy. *The Feminist Case against Bureaucracy*. Temple University Press, 1984.
- "Girl Powered Redefining the Face of STEM." RECF, 2025, recf.org/teams/for-participants/girl-powered/.
- "Girl Powered Robotics Workshop 2025." India Community Center, 2025, www.indiacc.org/girl-powered-robotics-workshop/.
- Gurung, Lina. "Feminist Standpoint Theory: Conceptualization and Utility." *Dhaulagiri Journal of Sociology and Anthropology*, vol. 14, 2021, pp. 106–115. doi:10.3126/dsaj.v14i0.27357.
- Koenig, Anne M., and Alice H. Eagly. "Evidence for the Social Role Theory of Stereotype Content: Observations of Groups' Roles Shape Stereotypes." *Journal of Personality and Social Psychology*, vol. 107, no. 3, 2014, pp. 371–392. American Psychological Association, <https://doi.org/10.1037/a0037215>.
- Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland. "Expectations of Brilliance Underlie Gender Distributions across Academic Disciplines." *Science*, vol. 347, no. 6219, 2015, pp. 262–265. American Association for the Advancement of Science, <http://dx.doi.org/10.1126/science.1261375>.
- Lin, Jessica. "Robotics Team Unveils Female Robots to Solve Gender Issues." *Upstream News*, 31 Mar. 2022, cvhsnews.org/9284/satire/robotics-team-unveils-female-robots-to-solve-gender-issues/.
- Maake, Tshepo B. "Reflecting on the Significance of the Feminist Standpoint Epistemology in Researching the Experiences of Black Gay Men Who Work in Heteronormative and Patriarchal Male-Dominated Workplaces." *Journal of Homosexuality*, 2025, pp. 1–18. Taylor & Francis Online, <https://doi.org/10.1080/00918369.2025.2501409>.
- McAlister, Faber. "Gender in Rhetorical Theory." *Oxford Research Encyclopedia of Communication*. 29 Nov. 2021, Oxford University Press. <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-75>.
- Miller, David I., Kyle M. Nolla, Alice H. Eagly, and David H. Uttal. "The Development of Children's Gender-Science Stereotypes: A Meta-Analysis of 5 Decades of U.S.

- Draw-a-Scientist Studies.” *Child Development*, vol. 89, no. 6, 2018, pp. 1943–1955. Wiley, <http://dx.doi.org/10.1111/cdev.13039>.
- Muhamad, Jessica, Tyler Harrison, and Fan Yang. “Organizational Communication.” *The International Encyclopedia of Media Literacy*, 2019, doi:10.4324/9780203710753-30.
- Parker, Patricia S. *Race, Gender, and Leadership: Re-Envisioning Organizational Leadership from the Perspectives of African American Women Executives*. Lawrence Erlbaum Associates, 2005.
- “About Us.” RECF, <https://recf.org/about-us/>. Accessed 12 Sept. 2025.
- Roberts, Sonia, and Alysson Light. “Collaborations, Not Competitions, Can Reduce Gender Disparities in Robotics.” *Feminist Pedagogy*, vol. 4, no. 2, 2023, article 8. California Polytechnic State University, <https://digitalcommons.calpoly.edu/feministpedagogy/vol4/iss2/8>.
- RobotEvents (2025). VEX V5 Robotics Competition. Web. <https://www.robotevents.com/robot-competitions/vex-robotics-competition>.
- “VEX V5 Robotics Competition.” Robot Events, www.robotevents.com/robot-competitions/vex-robotics-competition. Accessed 13 Sept. 2025.
- Rodriguez, Jenny K., and Elisabeth Anna Guenther. “Gendered Organization Theory.” *Oxford Research Encyclopedia of Business and Management*, Oxford University Press, 15 Sept. 2022. <https://doi.org/10.1093/acrefore/9780190224851.013.53>.
- Silver, Blake R. “Inequality in the Extracurriculum: How Class, Race, and Gender Shape College Involvement.” *Sociological Forum*, vol. 35, no. 4, 2020, pp. 1290–314. JSTOR, <https://www.jstor.org/stable/48642510>. Accessed 13 Sept. 2025.
- Sullivan, Amanda, and Marina U. Bers. “VEX Robotics Competitions: Gender Differences in Student Attitudes and Experiences.” *Journal of Information Technology Education: Research*, vol. 18, 2019, pp. 97–112, <https://doi.org/10.28945/4193>.
- Syed, Jawad, and Peter A. Murray. “A Cultural Feminist Approach Towards Managing Diversity in Top Management Teams.” *Equal Opportunities International*, vol. 27, no. 5, 27 June 2008, pp. 413–432, <https://doi.org/10.1108/02610150810882288>.
- Tannen, Deborah. *Talking from 9 to 5: Women and Men in the Workplace: Language, Sex and Power*. Avon Books, 1994.
- TeamTX. “Thoughts on Being a Girl in VEX?” VEX Forum, 2019, <https://www.vexforum.com/t/thoughts-on-being-a-girl-in-vex/70133>. Accessed 13 Sept. 2025.
- Tremmel, Manuela, and Ingrid Wahl. “Gender Stereotypes in Leadership: Analyzing the Content and Evaluation of Stereotypes about Typical, Male, and Female Leaders.” *Frontiers in Psychology*, vol. 14, 2023, article 1034258, <https://doi.org/10.3389/fpsyg.2023.1034258>.
- Weiner, Jennie M., and Laura J. Burton. “The Double Bind for Women: Exploring the Gendered Nature of Turnaround Leadership in a Principal Preparation Program.” *Harvard Educational Review*, vol. 86, no. 3, 2016, pp. 339–365, <https://doi.org/10.17763/1943-5045-86.3.339>.

A Study of AI's Impact on Education: Evaluating the Economic Sustainability and Demand for Traditional Tutoring Services In the US By Aaryan Kumar

Abstract

With the rapid development of artificial intelligence, the tutoring industry faces immense disruption, some of which might be impossible to rebound from. Students in the United States increasingly turn to ChatGPT and Khanmigo for support, reducing the interest in in-person tutoring traditions. While some previous research leans on student engagement and learning outcomes, how AI impacts the economic sustainability and demand for tutoring services in the US has not been examined. This paper explores how this disruption affects the sustainability and demand for tutoring services. Additionally, it displays the threat of AI to services that flourished in the past, while also describing the adaptability of tutoring an AI to create a more usable AI model. The study used a mixed-method approach, combining case studies, literature reviews, and surveys. With these approaches, the paper explores different factors such as personalization, adaptability, pricing, and trust, which are necessary to shift tutoring to AI. It debates the future of education by describing how AI can reshape and discuss how it can disrupt different long-lasting business models. The paper includes insights into market and student behavior, allowing it to educate others on how the education system is being reformed.

Keywords AI tutoring, Human tutoring, Personalization, Edtech

Introduction

As artificial intelligence develops in society, it inevitably alters our lives, including in the context of education. In the United States, tutoring jobs such as private academic coaching, test preparation, and homework help have been facing steady transformation. Companies, including Varsity Tutors, have been implementing tutoring services to assist students, yet new and accessible tutoring services utilizing AI have taken off. This technological shift raises urgent questions about the impact on the economic sustainability and demand on companies combating AI-tutoring. Understanding this technological shift is essential for companies adapting to the harsh tutoring climate. This study aims to answer the question “How are AI-based tutoring services affecting the economic sustainability and demand for traditional human-based tutoring services in U.S schools?”

The methodology leveraged in this paper consists of both qualitative and quantitative approaches. It uses a case study analysis that compares human-tutoring and AI-tutoring. Specifically, it determines the extent of disruption AI tutoring has caused on its more primitive counterpart. This is done by examining revenue trends, user engagement, and layoffs. Additionally, a survey was strategically implemented to capture the opinions of the public. Different samples of students were selected to ensure a diverse population, and efforts were made to achieve the largest sample size for the least margin of error. Lastly, the paper includes a literature review detailing academic papers, industry reports, and previous research to solidify

my knowledge. This assists in providing background information on the rise of AI and its impact on traditional tutoring services. These methods allow for an accurate assessment of the rapidly evolving tutoring landscape in the United States.

This study aligns with ongoing research regarding the intersection of AI and schooling by questioning the extent of the inverse relationship between the economic sustainability of human-based tutoring services before and after AI-tutoring services. It aims to discover whether increased use of AI tutoring tools like Khanmigo and ChatGPT directly correlate with the demand for human-based tutoring services. By combining case studies, surveys, and literature reviews, this research allows new information to be attributed to the field of technology concerning AI. It is essential due to our steadily growing society. This research presents pedagogical information and addresses the financial sustainability of tutoring services after the rise of artificial intelligence. The contribution this paper allows for new research is solely empirical.

While most past research on AI-tutoring focuses on learning outcomes, it fails to address the impact on the economic sustainability of human tutoring. Studies such as Létourneau et al. (2025) have glossed over the personalization and learning outcomes related to AI-based tools but fail to encapsulate how this new form of tutoring is reshaping the economics of human tutoring. Furthermore, studies often overlook students' perceptions of the cost and value of both human and AI-tutoring. With this knowledge, this empirical insight may reform the tutoring landscape.

The findings of this research are highly significant to tutoring businesses, teachers, and students alike in the US, where the technological adoption is very high. As we are on the uprise of artificial intelligence, it is natural for tasks to be increasingly adopted by AI. For example, tutoring companies may need to adapt their models to meet increasingly high student demands. On a similar note, schools can use this research to reform learning. Students may like AI to help with schoolwork more than a tutor, which can increase test scores.

As society develops technologically, it is substantially important to realize the changes AI brings not only to education, but also to our economy. This research helps traditional tutoring services thrive in our new economy, as their long-standing models face collapse in the face of modern, AI tutoring. Researching how they can adapt to change supports resilience in demanding educational ecosystems. Additionally, the study touches on digital equity; it mentions how AI can facilitate academic support.

Methodology

The methodology leveraged in this paper consists of both qualitative and quantitative approaches. The primary source of my research is by case study analysis, which seeks out comparisons between human and AI-tutoring. Specifically, the paper expounded on the trend of human tutoring before and after the ascent of AI use.

Furthermore, a survey was implemented to capture the opinions of the public. Different samples of students were selected to ensure a diverse population, and the survey strived for the largest sample size with the least margin of error. The survey was conducted using students from different grade levels and educational backgrounds. Questions honed in on students' opinions on the cost, effectiveness, and preferences of the two tutoring services. This information allows me to analyze and combine different data.

Lastly, the paper includes a literature review detailing academic papers, industry reports, and previous research to solidify my knowledge. This assists in providing background information on the rise of AI and its impact on traditional tutoring services. Combined, these methods allow for an accurate assessment of the rapidly evolving tutoring landscape in the United States.

Literature Review

This literature review aims to connect existing research on the impact of AI-tutoring on education with its economic implications, providing context on how previous research bridges the gaps between the transformation from traditional AI-tutoring. Combining existing research with ongoing research allows a more complete understanding of the topic.

Previous research on AI in schooling builds a foundation for how this research can be significant to the masses. Previous research surrounding AI usually entails how an ITS aids students, especially with personalization. Furthermore, most research relies on infrastructure, engagement, and learning outcomes. The review sources materials from academic papers, local reports, and surveys to ensure an unbiased result.

Table 1: Selected Literature

Study	Methods	Country	Factors	Findings
Létourneau et al. (2025)	Quantitative	USA	- Learning outcomes - Engagement - Design of the AI - Teacher's support	- Similar results were obtained. In some cases, ITS outperformed humans, but in others, it was the opposite. - ITS can personalize learning experiences. - The effectiveness of an ITS can also depend on the student's prior knowledge and cognitive ability. - ITS cannot replace a teacher, but can enhance overall learning.
Stritto et al.(2023)	Qualitative	USA	- School's willingness - Technological access - Trust in AI - Capabilities of the specific AI	- Many students were found not using AI for classes, but for personal and professional contexts - Most used it for brainstorming and generating content - Students at Oregon had mixed feelings about integrating their courses with AI
Arundel	Quantitative	USA	Student trust in	Using an AI-assisted tutor to curate their

(2024)			- AI - Reliance on technology - Tutor's readiness to adopt new methods	teaching style for each student. - Tutors who used the AI helped students gain 4% more on tests than students whose tutors didn't. - If the student answers a question incorrectly, the AI helps the tutor narrow down and ask similar questions to correct the student.
Keshtkar et al. (2024)	Mixed Methods	United States	- Student perception of tutoring services - The extent of personalization - Specific intent to use AI	- Students tend to have positive opinions regarding using AI as a tutor - Students envision tutoring using AI to be personalized for their strengths and weaknesses.
Danielle et al. (2024)	Quantitative	United States	- Learning outcomes - Income - Hybrid tutoring - engagement	- Hybrid human and AI models significantly improved learning and proficiency, especially for lower-achieving students. - AI support also reduced the total workload
Wang et al. (2024)	Quantitative	United States	- Tutor productivity - Cost - scalability	- Students who had tutors who were AI-assisted were 4% more likely to score higher than students with human tutors - AI helped tutors specify areas of work, and helped to correct mistakes faster, which increased overall speed
Kestin et al. (2024)	Quantitative	United States	- Learning efficiency - Time saved - engagement	- Students who used AI tutors learned twice as much in less time than students in an active learning class. - When using an AI tutor, students also felt more engaged and motivated than students in a class.
Berdik (2024)	Qualitative	United States	- Engagement - Hybrid - AI Vs Human	- AI can assist teaching by answering questions and providing easy access to resources, but it is unable to replace teachers fully. - AI enhances teachers and is not ready to replace them, especially in a complex classroom setting.

AI tutoring systems generally positively impact students (Létourneau (2025); Keshtkar, (2024). It is shown that students perceive AI as helpful to their academic goals, with the main positive as personalization. The AI model could personalize based on the strengths and weaknesses of the students, and in many cases was seen assisting teachers. Létourneau (2025)

was shown to focus on academic success and student personalization, while Keshtkar (2024) was seen to hone in on student perception of AI. Due to this, the conception can be made that personalization is not the only guarantee for adoption, yet other factors concerning user trust and perception also matter.

Studies like Arundel (2024) and Stritto (2023) explored how students and schools perceive AI more. Arundel focused on academic excellence and stated that students who used AI to prepare for tests generally had higher scores than students who didn't. Meanwhile, Stritto explained how students used AI for content creation, and a few used it for school. The majority of students were using it for professional contexts. Arundel's findings demonstrate AI correlating with enhanced grades, while Stritto's findings demonstrate hesitation to utilize the service.

More recent research (Danielle et al., (2024), shows that hybrid tutoring models can also improve proficiency, especially for lower-achieving students, where help is needed. This relates to Wang et al. (2024), who emphasized that AI can assist tutors in pinpointing mistakes at a faster rate and that AI also improves grade percentages. The studies correlate by showing that hybrid models also improve percentages and reduce the overall workload of human tutors and teachers.

AI tutors are more time-efficient, with students learning twice as much material in less time than in an average classroom (Kestin et al. 2024). However, Berdik (2024) underscored AI's inability to replace teachers in complex classroom settings, stressing its role as an enhancement tool rather than a replacement. This contrast shows that AI currently needs more work to fully take the role of a teacher, as teachers may have irreplaceable characteristics that help students thrive.

This study fits into the existing body of research as it adds an economic aspect to research about AI. The studies in the literature review mainly focus on academic performance or student perception, but economic trends are only considered in hindsight. Given AI's rapid prominence, it is necessary to understand the financial aspects it brings to other businesses. This paper brings unique research that can assist companies needing reform after the AI-tutoring surge.

Case Study

Chegg

Founded in 2005 to be a textbook rental site, Chegg quickly reformed due to society's transformation. Textbooks were going digital, and profits began to shrink. What's the solution? Start tutoring students. Chegg began to offer homework help, Q&A's, writing assistance, and exam prep resources. Chegg saw profits increase, as the stock rose from April 2020 to March 2020. From March 2020 to February 2021, Chegg's stock was at its peak. With millions of students working remotely, Chegg saw increased use of tutoring help substantially grow. The pandemic solidified it as one of the most prominent tutoring services. In late 2021, more students were transitioning to in-school learning, and Chegg saw its stock drop (Kruppa, 2024).



When ChatGPT launched, the Former CEO of Chegg, Dan Rosensweig, publicly acknowledged ChatGPT's impact on his service. "Since March, we have seen a significant spike in student interest in ChatGPT. We now believe it's affecting our new customer growth rate." Stock dropped around 48% in one day, which erased billions in valuation (Novak, 2023). Meanwhile, student behavior also mimicked this shift. Growth Memo surveyed that only 30% of college students planned to use Chegg, down by an additional 8%. At the same time, 62% of students planned to use ChatGPT instead (Kruppa, 2024).

In April 2023, they unveiled their new AI. I have to combat the uprising of CheggMate. Partnered with OpenAI's GPT-4, CheggMate was designed to integrate personalized learning pathways and specifically tailored quizzes into its already comprehensive 150,000 human tutors. Rosensweig emphasized that combining Chegg's knowledge with GPT-4 would make the most potent "study companion" (Chegg, n.d.). The company's primary goal is to use its already robust library of answers to create a new hybrid that doesn't have hallucinations, like ChatGPT. When polled, testers of the latest model showed that 77% were excited about the newest AI model, but 85% were more eager when they learned it would be backed with human verification (Ghosh, 2023). Some named the announcement of the hybrid as too late, as stocks only rebounded 7%, after a significant 40-50% loss (Online Education, 2023).

Reflecting, the once prominent tutoring service Chegg was obsolete in the face of ChatGPT. Its experience with AI highlights the threat that AI poses to tutoring services. The student migration displays how they find instant, personalized, and free responses more valuable. The surveys showed Chegg's steep decline, as more students tended to leave the service for ChatGPT. Despite proclaiming the launch of CheggMate, some critics think it is too late for the service to prosper. This case underscores how AI is reforming the economic tutoring landscape, translating to the scarcity of relevancy.

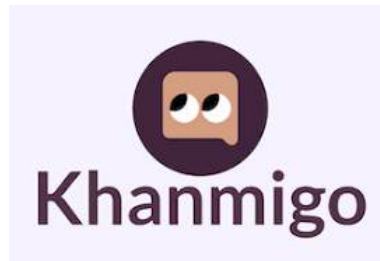
Khanmigo

Khan Academy is a nonprofit organization founded by Sal Khan in 2006. The platform is meant to have tutoring videos and exercises available worldwide. It utilizes mastery to display one's skill on the material, and has content from K-12 subjects and standardized testing. Before using AI, Khan Academy was a beacon of free and self-guided tutoring (Boryga, 2024). Khanmigo was founded for the sole purpose of revolutionizing education and providing personalized tutoring. This is much cheaper for \$4 a month than the other AI tutoring competitors (Axios AI+, n.d.).

Khanmigo is an AI-driven personalized tutor built with GPT-4. It directly integrates with Khan Academy's wide selection of videos and offers a wide range of help, including

problem-solving and step-by-step instruction. For educators, Khanmigo more efficiently creates lesson plans and handles administrative tasks, allowing more time for teachers to engage with students. Unlike chatbots like ChatGPT, it has a direct source for every one of its answers (Khan Academy, n.d.). Khan Academy clarifies that Khanmigo complements human tutors, as it doesn't directly give students answers. Instead, it focuses on helping the student learn the process and make advancements, leading them to finish the problem. Sal emphasizes that the new AI should not be used as an answer key, but as a tool to assist learning. Additionally, he mentions that AI should not be used as a replacement for educators, but instead to assist them with teaching (Perkel, 2025).

Khan Academy's AI Model, Khanmigo, shows how AI can coexist with teachers and tutors and enhance human tutoring. By decreasing the load off teachers, the AI makes it easier for teachers and tutors alike to work with students. Khanmigo supports a framework where teachers and AI can work together to amplify the effect on the student. This study shows that when used thoughtfully, AI tutoring can positively affect the human tutoring economy, and instead of replacing tutors, it redefines tutoring roles (Harvey Mudd, 2025).



Varsity Tutors

Varsity Tutors is a tutoring service that offers live online tutoring for test prep and academic courses. Built for students, the company emphasizes one-on-one and small group instruction. In 2025, the company launched its Live + AI platform. This integrates AI tools into live tutoring. Like hybrid tutoring, the AI is used to lessen employee workload, automatically matching students to tutors based on learning styles. Additionally, the platform automates lesson prep and displays analytics during sessions (Lawler, 2025). Live + AI ensures alignment with mandates by offering scalable and policy-compliant tutoring to schools (Varsity Tutors, n.d.).



Over 1,000 school districts have partnered with Varsity Tutors, significantly reducing teacher workload (Stone, 2025). Varsity Tutors is just one of the many cases where hybrid tutoring enhances human instruction instead of replacing it. As AI tools continue to be integrated into society, platforms utilizing AI and Human instruction may be necessary to maintain relevancy.

Varsity tutor's case restates a trend in society: companies that integrate AI into their tutoring service remain sustainable. It uses a hybrid method of AI and traditional tutoring, which allows it to remain integrated into society. With this, it is possible to make tutoring more accessible while increasing staff efficiency. As tutoring services struggle to adapt, Varsity Tutors differs and shows the others an example of effectively adapting to AI pressure.

Layoffs

After the uprising of AI, several edtech and tutoring companies in the U.S. have been making significant layoffs. Chegg, a tutoring company, laid off 22% of its workforce in May 2025 due to ChatGPT's impact on the service (Reuters, 2025). From January 2023 to December 2024, there were over 40,000 layoffs in the broader sector, which decreased the overall number of employees by 18% (DigitalDefynd, 2023).

The layoffs are linked to pandemic market corrections and AI disruption. During COVID-19, companies like Chegg saw a surge of users, which spiked their revenue. When learning in schools, companies saw decreases in revenue. At the same time, the rise of generative AI directly influenced the edtech industry. Offering free, instant, and personalized help, this is precisely what students were looking for. Subsequently, the funding for edtech decreased: \$17.3 billion in 2021 to \$3 billion in 2024 (Criddle, 2024). With less funding and overall sales decreasing, companies needed to reformat. For example, Chegg closed several offices to save "between \$45 million and \$55 million in 2025" (Vaziri, 2025). While AI isn't the entire catalyst for layoffs, it has accelerated the need to manipulate edtech companies. The layoffs and restructuring of companies suggest that AI is steadily replacing the need for human tutors or, in other cases, supporting them. In one instance, Paper, an edtech startup, has also been undergoing cutoffs as it eliminates many tutor roles. Declaring COVID-19 the leading cause of their loss of funding, the edtech industry, facing staff cuts and funding freezes, illustrates the instability of jobs for tutors.

Comparative Analysis

The case study highlights the economy's vulnerability of traditional tutoring models, and how AI models have changed them. As students prefer the new and personalized AI solutions, companies that rely on human tutors are seeing reduced income and financial instability, resulting in layoffs. This underscores the need to properly examine the economic sustainability and demand shifts in tutoring services, as companies in the past have been rendered unstable due to their new and instant AI counterpart. See Table 2 for the comparative analysis across AI tutoring companies' case studies.

Table 2: Comparative Analysis Across AI tutoring companies in the US

Case / Company	Model	AI Impact	Human interaction	Economic Impact
Varsity Tutors	1:1 tutoring, Hybrid	AI matches students, plans lessons, etc.	Tutoring is still stable, actively teaching	Educator workload reduced
Chegg	Textbook, online tutoring	AI chatbot integration	ChatGPT caused a significant disruption	Stock dropped about 50%, and layoffs took place
Paper	Chat-based live tutoring	AI assists with tasks	Tutors still assisting with live responses	Model balances AI with human chat
Layoffs	Varied (primarily online)	AI replaced tasks previously done by tutors	Tutor displacement	Job losses among private companies
Khanmigo	Free education platform, paid AI add-on	AI add-on supports recorded videos and practice	AI helps the tutoring instruction	Stable, higher efficiency

With the case studies of Chegg, Varsity Tutors, Khan Academy, and Layoffs in the edtech industry, one thing becomes clear: companies that fail to adapt to the new and emerging AI face demand and financial performance declines. On the other hand, companies like Khan Academy, which have adapted to AI, tend to experience sustainability. Varsity Tutors maintains stability, as they have utilized AI for scheduling and matching students, while also still relying on tutors. Layoffs in the industry demonstrate that companies that didn't change their business models for AI were prone to post-pandemic market disruptions. Adaptability and AI integration were key factors for determining sustainability in the tutoring industry. The demand for traditional tutoring has not disappeared, but has more likely been reformed. For example, students now prioritize instant access, cost, and convenience; AI tutoring outperforms traditional tutoring. Due to its instant responses, surveys show that 62% of college students prefer ChatGPT over Chegg. Additionally, the shift to AI is not just based on convenience, but also on personalization. Khanmigo demonstrates the need for human verification and AI to create a feasible tutoring business, suggesting that hybrid models, which combine AI efficiency with human expertise, may be necessary in our developing society.

The case studies show that traditional tutoring services may undergo immense challenges unless they adapt to AI technology. AI is replacing tutoring and reinventing what is in its best interest. While Chegg's late action on collaborating with AI shows the risk, Khanmigo's success shows that AI can enhance tutors rather than replace them. This provides insight into my research question by highlighting that adaptation is needed to maintain the economic sustainability of traditional tutoring rather than compete with AI.

Survey

For a better understanding of the impact of AI on tutoring services, a survey was distributed online with Google Forms during a week in July 2025. The survey had responses from 100 participants and included all multiple-choice questions. The study's respondents were students, educators, tutors, and older people. This was chosen to fully encapsulate the opinions of AI from a wide age range. The goal of the survey was to gather data on how people are adapting to AI tools, and whether it has changed their or someone they know – reliance on human tutoring.

The questions in the survey focused on the frequency of AI use, preferred platforms, and opinions on cost. They honed in on whether they believed AI was replacing the need for human tutors, and how speed, personalization, and price were reflected in their answers. The primary goal focuses on stakeholders' widespread choices to stray from or rely on human tutoring. Other questions relied on opinions on the future of AI.

Ethical considerations were taken seriously in the survey. It was anonymous, meaning no data was extracted except for the responses, and no information was collected. All participants answered voluntarily and honestly and acknowledged that their responses were only used for academic purposes. The questions in the survey were worded objectively to avoid bias in participant responses. The survey results capture the evolving tutoring landscape. By capturing both educators and students, the data helps form an accurate response about attitudes to AI services and these educators' trust, perceptions about cost, and demand for human-tutoring services.

Table 2: Survey Questionnaire and Responses

Hi, my name is Aaryan Kumar. I am a rising senior working on a research paper project. With the results of this survey, I aim to understand the reasons for how AI tutoring services affect traditional tutoring services. The goal is to understand the reasons behind shifts in tutoring practices driven by technological advancements. The collected data will be allocated to serve research to answer the question of “How are AI based tutoring services affecting the economic sustainability and demand for traditional human-based tutoring services in U.S schools?” I will ensure that your responses are confidential and only used for educational purposes. Thank you!		
Survey Questions	Response Options	% Response
1. Have you ever used an AI tutoring tool? (Ex. ChatGPT, Khanmigo)	1. No 2. Yes	• 2% • 98%
2. How often are you using AI for academic help?	1. Weekly 2. Monthly 3. Annually	• 75% • 25%
3. When was the last time you worked with a human tutor?	1. 1-2 Months ago 2. 3-4 Months ago 3. 5-6 Months ago 4. Over 6 Months 5. Never	• 20% • 18% • 16% • 26% • 20%
4. Which type of tutoring do you	1. Human tutoring	• 5%

prefer?	- AI tutoring - Mix of both Human & AI tutoring - None	48% 40% 7%
5. I feel like an AI tutor would help me better than a human tutor	- Strongly Disagree - Disagree - Neutral - Agree - Strongly Agree	2% 25% 28% 18% 27%
6. I would feel more confident about improving my grades using a human tutor	- Strongly Disagree - Disagree - Neutral - Agree - Strongly Agree	1% 30% 18% 30% 21%
7. A human tutor would be more personalized for me than an AI tutor.	- Strongly Disagree - Disagree - Neutral - Agree - Strongly Agree	0% 26% 22% 22% 29%
8. Which one would you think is more affordable?	- Human tutoring - AI tutoring - None	2% 97% 1%
9. In which subjects do you use tutoring for?	- Math - History - Science - English - Foreign language - Test prep (SAT, ACT) - Other - None	45% 42% 57% 36% 37% 53% 55% 1%
10. What are your main reasons for AI tutoring?	24/7 availability - Instant doubt resolution - Low emotional connection - Adaptive explanations - Limited critical feedback - Creativity boosters (e.g., brainstorming prompts) - Mistrust in complex tasks - Fun, gamified formats	41% 40% 41% 32% 33% 40% 32% 33% 1%
11. Which of the following factors might lead to challenges in accessibility and affordability for AI tutoring?	- Device/internet divide - Paid tool barriers - Awareness gaps (rural/low-tech areas) - Free vs paid AI access - Language limitations	53% 43% 54% 49% 42% 42%

	6. High perceived value (urban/tech-savvy users)	
12. What factor seems most important for making AI tutoring more adaptable?	1. AI feedback loops 2. Hybrid learning models 3. Tech training for tutors 4. Customization flexibility 5. Platform fatigue 6. Switching resistance	• 13% • 19% • 19% • 14% • 20% • 15%
13. What do you think has been the most important economic impact of AI tutoring?	1. Decline in private tutor hiring 2. Revenue shifts (human → AI firms) 3. Edtech investment surge 4. Tutor job losses 5. Subscription-based pricing models 6. Emergence of new AI-related roles	• 19% • 17% • 11% • 22% • 10% • 21%
14. What outcomes do you expect for the future of AI tutoring?	1. Higher short-term test gains 2. Improved engagement (e.g., shy learners) 3. Limited deep learning 4. Weak metacognitive growth 5. Student tool preference divergence 6. Increased independence, reduced scaffolding	• 52% • 46% • 41% • 50% • 47% • 41%
Age	• Below 18 • 18-25 yrs • 25-35 yrs • 35-45 yrs • 45-60 yrs • 60 and above	• 23% • 17% • 12% • 14% • 17% • 17%
Gender	• Male • Female • Other • Prefer not to say	• 42% • 29% • 5% • 24%

Analysis of the survey

The survey results demonstrate patterns in how AI is transforming the tutoring landscape. A diverse group easily reflected educators' and tutors' opinions. Three themes stood out from others: affordability, preference, and efficiency. Most respondents favored AI tutoring or a hybrid of AI and human tutoring. This indicates a major societal shift in the perspective on tutoring, with many people favoring 24/7 availability, personalized explanations, and instant feedback over human interaction. These features underscore the importance of flexibility in the steadily transforming edtech industry. Affordability plays a substantial role in favoring AI tutoring over human tutoring. 97% of participants thought tutoring with AI would be more affordable than tutoring with a human being. Since traditional tutoring can be expensive or inaccessible, AI can

be used as a rebound by offering wide-access and free services. This assists communities that are under-resourced and lack willing educators. Even though AI was seen as efficient and accessible, many felt human tutors still had an edge over AI tutoring. For example, 51% of participants agreed that human tutoring would improve grades, while 49% believed AI tutoring would improve grades. Additionally, 51% of respondents also thought that a human tutor would be more personalized for them. This shows that AI tutoring is not a complete replacement for human tutoring, as the latter still has advantages that some students may look for when studying.

In sum, the data shows that stakeholders are open to AI tools in education, primarily because of their convenience and affordability. Consequently, for AI to truly encapsulate all the public's needs, it is essential to improve personalization and trust. Students and educators still believe that human tutoring would increase their knowledge further, which shows that there are more areas for AI tutoring to cover in the future.

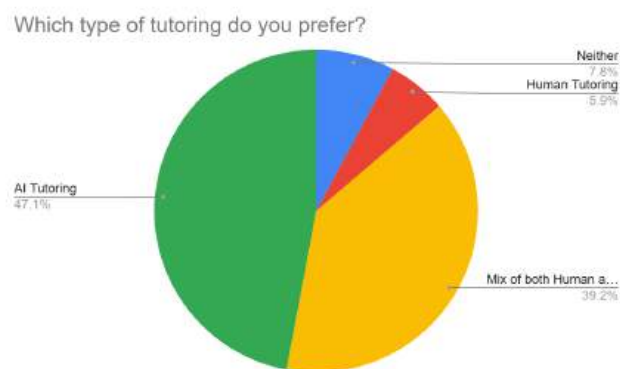


Figure 1: Type of Tutoring

Figure 1 chart represents responses to the question “Which type of tutoring do you prefer?” Choices include AI Tutoring, Human Tutoring, Mix of Human and AI Tutoring (Hybrid), and Neither. The most preferred option out of the four was AI Tutoring (47.1%), followed by Hybrid (39.2%), then Human Tutoring (5.9%), and lastly Neither (7.8%). This demonstrates that the leading choice is AI tutoring, with many participants also leaning towards a blended option with a mix of the two.

The least picked option was Human Tutoring. This suggests an unseen decline of human tutoring, with participants opting for no tutoring over human tutoring. The high percentage of stakeholders opting for Hybrid tutoring shows that they might still value human interaction, but only when combined with AI benefits such as speed, adaptability, and convenience. The data supports that AI is not just an uprising tool; it is taking over the academic industry. It displays attitudes about changing values, such as more reflective opinions about social trends. All of these tie into the question about AI’s role in changing the economic sustainability of traditional tutoring.

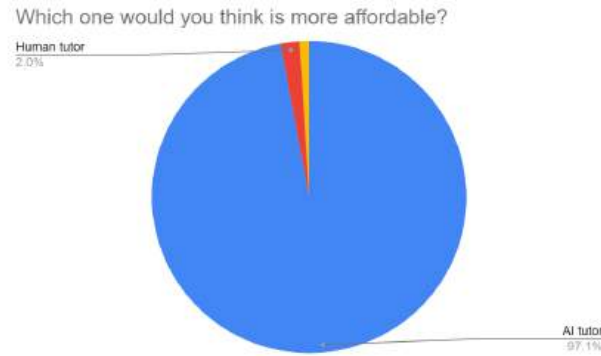


Figure 2: Affordability

This pie chart addresses the question “Which one would you think is more affordable?” A staggering portion of responses favor AI tutoring, with 97.1%. Only 2% think Human tutoring is more affordable. This overwhelming result rates affordability as one of the driving factors for the AI industry to take over edtech. This is a key concern for students and educators alike, especially in our technology-focused society. Accordingly, it is essential to acknowledge that this may not be a preference. Access can play a huge role in affordability, as some people cannot access human tutoring and have to opt for the free and instant AI tutoring.

These two graphs demonstrate that affordability is among the most critical factors behind the preference for AI tutoring. While users still value human interaction, survey responses show that the benefits of AI outweigh the preference for traditional tutoring services. These findings reinforce the argument that AI is reforming the tutoring landscape and better aligning with user demands about affordability.

Discussion

This framework was developed after a survey, literature review, and multiple case studies. Each box in the diagram represents a key recurring theme that emerged when referring to different methods. The student perceptions were developed from literature reviews and later added to the surveys. The center of the framework shows the cause and effect with accessibility, perception, and adaptability leading to economic impact and outcomes. This also indicates that AI tutoring is not just one topic, but is a branch of different issues all connecting to one idea, with set outcomes built on it.

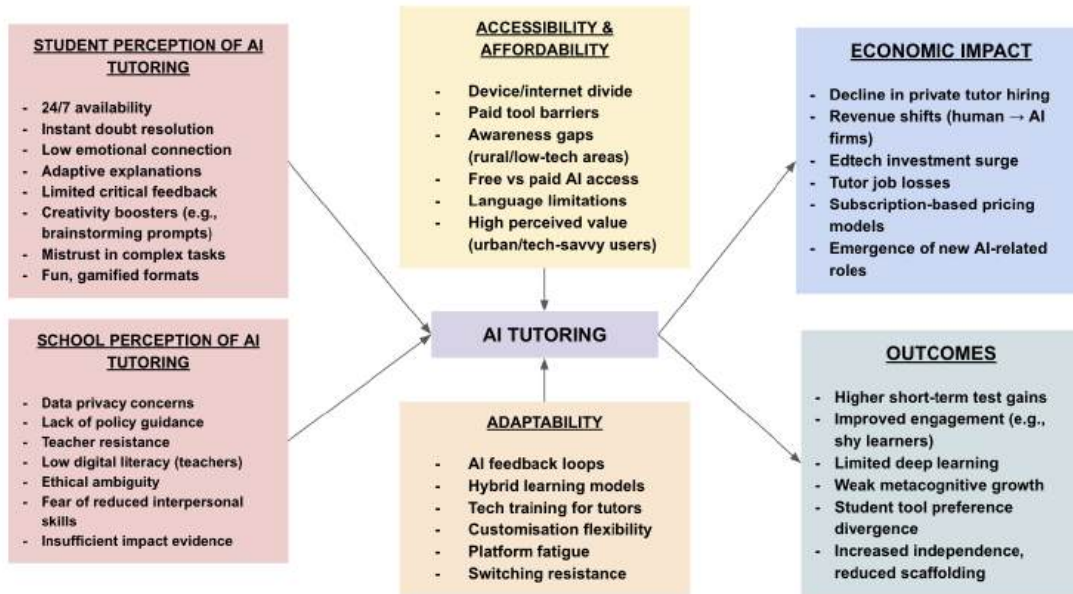


Figure 1: Process Framework for Factors affecting AI tutoring and the outcomes

This paper introduces a framework categorizing AI tutoring research into student and school perception, accessibility and affordability, adaptability, outcomes, and economic impact. Derived from existing literature, the framework shows how these components influence outcomes. Findings indicate AI is transforming, not replacing, human tutoring, with students embracing it while some educators are hesitant. Hybrid tutoring, combining human and AI elements, proves most effective. However, AI's disruptive economic impact, evidenced by layoffs, is also highlighted. The framework aims to guide edtech companies, teachers, and tutors in AI adoption, for instance, by helping teachers integrate AI to improve test scores. A deeper understanding of these categories can help companies develop sustainable AI tutoring models.

Students tend to view AI tutoring positively due to its instant responses and infinite availability. It has adaptive responses and is used for generational content. Consequently, they also show concern for low emotional connection and limited critical feedback. These factors are inevitable, as no human guides you through problems or practice. This perception directly affects trust and usage, as it influences tutoring outcomes. As students are the majority of the academic AI user base, their opinion is the most valuable. School administrators are hesitant to adopt AI due to concerns about data privacy, lack of policy, and potential reduction in skill development. Teachers also express low digital literacy. Despite AI's efficiency, barriers like limited internet, paid tools, and awareness gaps hinder accessibility, impacting global economic scaling. AI tutoring's success relies on adaptability, including feedback loops and customization, but requires tutor training. While AI offers short-term gains like increased test scores and engagement, it shows weak metacognitive growth and limited deep learning. These outcomes reflect school and student perceptions, adaptability, and affordability. With job losses, declining private tutor hiring, and revenue shifts, AI tutoring has reformed the economic landscape. The service is taking away

from traditional tutoring services, sometimes bankrupting them. Adaptation is needed for survival, especially economically. New AI roles and subscription models further entail how disruptive the feature is for the economy. This is due to factors such as student perception, affordability, and accessibility, as these shape whether AI gets funding or human tutoring goes out of the picture.

Limitations and Future Directions

This study acknowledges its limitations, crucial for future research. While offering insight into the tutoring industry's changes, it's constrained by data, scope, and methodology. Firstly, the survey-based design, though collecting 100 responses across age groups, is small compared to the U.S. tutoring population. The online format may have biased respondents towards AI. It also excluded business owners, bystanders, and parents, who are arguably knowledgeable stakeholders, due to time constraints. Future research should use mixed methods, including interviews, larger groups, and broader data collection. Secondly, the study focuses solely on AI tutoring in the U.S., specifically K-12 and college sectors. This limits global applicability and excludes international systems where AI varies significantly. The research also primarily covers popular platforms like Chegg, Varsity Tutors, and Khan Academy, potentially not reflecting trends in smaller businesses. Future research should compare across countries to understand diverse AI impacts, exploring new regions for economic insights and unique advantages. Finally, self-reported survey responses introduce limitations, such as social desirability bias or keyword confusion: the survey prioritized preferences, opinions, and future choices over precise data. Future research could incorporate diverse data sources, such as school analytics or student performance data from teachers, along with institutional input for more accurate, concrete data. Comparing past and present trends would allow for a better evaluation of AI tutoring's long-term influence.

Conclusion

This study aimed to investigate how artificial intelligence is reshaping the tutoring industry and how the economic sustainability and demand for traditional tutoring services have changed over time. This was conducted with a lens on student and school perception, affordability, accessibility, and adaptability. Through a mixed-methods study approach, which included a survey, literature review, and case study, the research focused on how AI services are changing the edtech industry.

From the study, the findings display how the way tutoring is perceived has changed. AI tutors are not taking over human tutoring but changing how it functions. Students from the survey and literature review showed that they preferred AI tutoring and hybrid tutoring. They highlighted preference for personalization, affordability, and overall convenience as the main factors for their choice. Case studies such as Varsity Tutors and Chegg helped convey the different economic impacts. For Chegg, AI completely ruined their financial sustainability and led to their downfall. Varsity Tutors opened a new perspective, showing a more adaptable model that includes AI, which helps in scheduling, formatting, and matching. Their hybrid model

demonstrates the duality of perception and shows that adapting to AI is crucial. Stakeholders from the survey show that while some users trust AI for academic help, human tutoring still offers greater personalization.

The findings from this research show that the tutoring industry is transforming. The rise of hybrid tutoring underscores that AI is most effective when paired with human instruction rather than replacing it. Platforms that have incorporated AI into their models, such as Khan Academy and Varsity Tutors, have seen unprecedented benefits. Meanwhile, companies that fail to adapt to the AI uprising are shown to face economic decline, as seen in Chegg. However, the study was also limited as it only covers the United States, so it may not fully encapsulate global tutoring trends. Furthermore, the scope of K-12 and early college-level tutoring models may also limit the general data, as the study only covers a portion of academic levels. Another limitation was the self-reported responses, which may include bias due to societal trends or socially desirable answers.

Further research can widen the scope of economic sustainability and demand and experiment globally or in a country with an economy much different from the U.S. Studies similar to this can provide better insight into how tutoring has changed over time, while research comparing different tutoring models, such as hybrid, AI, and human, can measure total efficiency. This research offers information for remodeling tutoring services to adapt to AI. In the edtech industry, companies that tried to withstand AI drowned, and companies that adapted survived. They should consider hybrid tutoring platforms that combine AI efficiency with human experience and personalization. It also shows how training in schools is needed to bridge the gap in adaptability, and tutors should reconsider whether their income would be stable based on AI interference.

In conclusion, AI is not taking over human tutoring or eliminating it; it is evolving it. As AI use increases and students lean towards instant, 24/7 responses, hybrid tutoring will likely be the center of it all. This study addresses both challenges and opportunities and paves the way for new edtech models to take steps toward adaptation.

Works Cited

- Létourneau, Angélique, et al. "A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education." *npj Science of Learning* 10.1 (2025): 1-13.
- Stritto, Mary Ellen Dello, Greta R. Underhill, and Naomi R. Aguiar. "Online Students' Perceptions of Generative AI." (2024).
- Arundel, Kara. "How AI Can Improve Tutor Effectiveness." *K-12 Dive*, 7 Oct. 2024, www.k12dive.com/news/ai-tutor-effectiveness-stanford-university/728980.
- Keshtkar, Fazel, et al. "AI Tutor: Students' Perceptions and Expectations of AI-Driven Tutoring Systems: A Survey-Based Investigation." *The International FLAIRS Conference Proceedings*. Vol. 37. 2024.
- Al-Sibai, Noor. "Chegg's Stock Crashes as ChatGPT Steamrolls Its Business." *Futurism*, 3 May 2023, <https://futurism.com/the-byte/chegg-stock-chatgpt>
- Kruppa, M. (2024, November 9). How ChatGPT brought down an online education giant. *WSJ*. <https://www.wsj.com/tech/ai/how-chatgpt-brought-down-an-online-education-giant-200b4ff2>
- Indig, Kevin. "CheggMate." *Growth Memo*, 3 Mar. 2025, www.growth-memo.com/p/cheggmate
- Maxwell, Thomas. "Chegg Is on Its Last Legs After ChatGPT Sent Its Stock Down 99%." *Gizmodo*, 25 Feb. 2025, gizmodo.com/chegg-is-on-its-last-legs-after-chatgpt-sent-its-stock-down-99-2000522585
- "Chegg Launches CheggMate AI Companion, Built With GPT-4." *TradingView*, 17 Apr. 2023, www.tradingview.com/news/benzinga%3A57120615b094b%3A0-chegg-launches-chegg-mate-ai-companion-built-with-gpt-4
- Meet Khanmigo: Khan Academy's AI-powered Teaching Assistant and Tutor. www.khanmigo.ai
- Perkel, Sarah. "Khan Academy CEO Predicts AI in the Classroom Will Be Like 5 'Amazing Graduate Students' Assisting Teachers." *Business Insider*, 25 June 2025, www.businessinsider.com/khan-academy-ceo-ai-classroom-teacher-aid-not-replacement-2025-6.
- Thomas, Danielle R., et al. "Improving student learning with hybrid human-AI tutoring: A three-study quasi-experimental investigation." *Proceedings of the 14th Learning Analytics and Knowledge Conference*. 2024.
- Wang, R. E. *, Ribeiro, A. T., Robinson, C. D., Loeb, S., Dora Demszky, & Stanford University. (2025). Tutor CoPilot: A Human-AI approach for scaling Real-Time expertise [Journal-article]. <https://arxiv.org/pdf/2410.03017>
- Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2024). AI tutoring outperforms active learning. *Research Square* (Research Square). <https://doi.org/10.21203/rs.3.rs-4243877/v1>
- Berdik, C. (2024, July 10). AI can't replace teaching, but it can make it better. *WIRED*. <https://www.wired.com/story/what-aspects-of-teaching-should-remain-human/>
- Novak, J. (2023, December 7). Chegg Stock Gets Crushed by ChatGPT, Causing Shares to Fall 48% – Is it Time to Sell? *VectorVest*. <https://www.vectorvest.com/blog/hot-stocks/chegg-stock-gets-crushed-by-chatgpt/>
- Chegg announces CheggMate, the new AI companion, built with GPT-4. (n.d.). <https://investor.chegg.com/Press-Releases/press-release-details/2023/Chegg-announces-CheggMate-the-new-AI-companion-built-with-GPT-4/default.aspx>

- Ghosh, S. (2023, April 19). Introducing CheggMate: The AI-Powered Study Companion Built with GPT-4. Medium.
<https://medium.com/the-arttech-future/introducing-cheggmate-the-ai-powered-study-companion-built-with-gpt-4-25d88111b0aa>
- ChatGPT crashes Chegg's stock 50 percent in one day. (2023, August 25).
<https://www.onlineeducation.com/features/chatgpt-crashes-cheggs-stock>
- Boryga, A. (2024, November 27). How AI will impact the future of teaching—A conversation with Sal Khan. Edutopia.
<https://www.edutopia.org/article/how-ai-will-impact-the-future-of-teaching-a-conversation-with-sal-khan/>
- Axios AI+. (n.d.). Axios. <https://www.axios.com/newsletters/axios-ai-plus-a8cc44b0-9492-11ef-b253-f943c4bfa123>
- Pillay, T. (2024, September 5). Kristen DiCerbo. TIME.
<https://time.com/7012801/kristen-dicerbo/>
- Why AI won't destroy education | Fall/Winter 2024–2025 | Harvey Mudd College Magazine. (n.d.). Harvey Mudd College Magazine.
<https://magazine.hmc.edu/fall-winter-24-25/why-ai-wont-destroy-education/>
- Hoddie, A. (2024, September 6). Varsity Tutors replaces paper tutoring. M-A Chronicle.
<https://machronicle.com/varsity-tutors-replaces-paper-tutoring>
- Lawler, R. (2025, May 1). Varsity Tutors launches Live+ AI Platform for schools — EdTech Innovation Hub. EdTech Innovation Hub.
<https://www.edtechinnovationhub.com/news/varsity-tutors-launch-live-ai-platform-for-schools>
- Varsity Tutors signs White House pledge to advance AI education for America's youth. (n.d.).
<https://investors.nerdy.com/news/news-details/2025/Varsity-Tutors-Signs-White-House-Pledge-to-Advance-AI-Education-for-Americas-Youth/default.aspx>
- Reuters. (2025, May 12). Chegg to lay off 22% of workforce as AI tools shake up edtech industry. Reuters.
<https://www.reuters.com/world/americas/chegg-lay-off-22-workforce-ai-tools-shake-up-edtech-industry-2025-05-12/>
- DigitalDefynd, T. (2025, June 30). Is the EdTech industry dying? [2025] - DigitalDefynd. DigitalDefynd. <https://digitaldefynd.com/IQ/is-the-edtech-industry-dying/>
- Criddle, C. (2024, December 24). Investment in online education groups plummets following the rise of AI. Financial Times.
<https://www.ft.com/content/54e8d249-8b95-44df-8bb4-c48ea20c7857>
- Vaziri, A. (2025, May 13). Chegg cuts 248 jobs, closes offices as AI rivals erode business. San Francisco Chronicle.
<https://www.sfchronicle.com/tech/article/chegg-layoffs-ai-competition-20325103.php>
- Mandel, C., & Mandel, C. (2023, September 15). Paper cuts more staff in second round of layoffs since August | BetaKit. BetaKit.
<https://betakit.com/paper-cuts-more-staff-in-second-round-of-layoffs>

Breaking the Biofilm Barrier: Tackling Antibiotic-Resistant Infections with Gene Editing, Quorum Sensing, and AI-Driven Solutions By Finneas Murphy

Abstract

Antibiotic resistance is an escalating global health crisis, threatening the efficacy of modern medicine and putting countless lives at risk. Each year, drug-resistant bacterial infections are responsible for at least 35,000 deaths in the United States alone, a number that continues to climb [1]. This crisis is driven by both natural bacterial evolution and the limitations of our current infection-fighting tools [2]. A major contributor to resistance is the formation of biofilms, structured communities of bacteria that adhere to surfaces and produce a protective extracellular matrix, making them extremely difficult to eradicate with conventional antibiotics [4][5]. Biofilms are especially problematic in hospital environments, where they frequently develop on catheters, ventilators, and prosthetic devices [4].

In addition to biofilms, bacteria also rapidly develop and spread resistance genes through horizontal gene transfer (HGT), a mechanism that allows genetic material to be exchanged between organisms via plasmids, transformation, or transduction [7–9]. This process accelerates resistance across bacterial populations and has led to the emergence of multidrug-resistant organisms such as CRE (carbapenem-resistant Enterobacteriaceae) and MRSA (methicillin-resistant *Staphylococcus aureus*) [2][3].

To confront these urgent challenges, researchers are developing novel and multidisciplinary solutions. Gene editing tools, particularly CRISPR-Cas systems, offer the possibility to selectively target and eliminate resistance genes [11–13]. Quorum-sensing inhibitors aim to block bacterial communication and prevent biofilm formation [14–17], while artificial intelligence is revolutionizing drug discovery by predicting resistance patterns and identifying new antimicrobial compounds [18–19]. Though each strategy has limitations, together they represent the foundation of a promising future in which antibiotic resistance can be tackled with precision, innovation, and resilience.

Keywords Antibiotic Resistance, CRISPR, Biofilm, Horizontal Gene Transfer, Quorum Sensing, Artificial Intelligence, Bacterial Evolution, Infection Control, Novel Therapies

Methods

A systematic literature review was performed to compile and analyze current research on antibiotic resistance mechanisms, including biofilm formation, horizontal gene transfer, and the limited development of new antibiotics [1–3]. The review also explored emerging therapeutic approaches such as gene editing (e.g., CRISPR-Cas), quorum-sensing inhibition, and artificial intelligence-based drug discovery [11–13,14–19]. Peer-reviewed studies and review articles were sourced primarily from PubMed, Google Scholar, and related biomedical databases.

1. Biofilm Formation and Antibiotic Resistance

1.1 What Are Biofilms?

Biofilms are complex, multicellular bacterial communities encased in a self-produced extracellular polymeric substance (EPS), which anchors them to surfaces such as tissues, implants, and medical devices [4][5]. This EPS matrix is composed primarily of polysaccharides, proteins, and lipids, forming a dense and protective barrier against external threats [4]. Unlike planktonic (free-floating) bacteria, cells within a biofilm can enter dormant states or exhibit low metabolic activity, making them highly tolerant to antibiotics and immune responses [4][5]. Research shows that bacteria within biofilms can be up to 1,000 times more resistant to antibiotics than their planktonic counterparts [4][5]. This makes biofilms one of the most formidable obstacles in the treatment of persistent infections, particularly in healthcare settings [1][4].

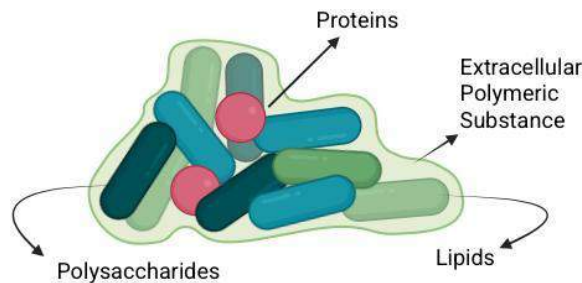


Figure #1: Structure of bacteria film

Structure of a bacterial biofilm. The extracellular polymeric substance (EPS) matrix covers cells in a protective barrier that restricts antibiotic penetration. This structural defense makes biofilm-associated infections up to 1,000 times more resistant to treatment than free-floating bacteria.

1.2 Why Biofilms Are a Clinical Threat

Biofilms are implicated in over 80% of chronic bacterial infections [5]. They are especially prevalent in hospitals, where they form on invasive devices such as urinary catheters, endotracheal tubes, pacemakers, and prosthetic joints [4][5]. Once established, biofilms act as reservoirs of infection, with bacteria capable of detaching and spreading through the bloodstream, leading to secondary conditions like sepsis or endocarditis [4][5]. Treatment often requires long-term antibiotic use and, in many cases, surgical removal of the infected device [4]. These infections increase healthcare costs, prolong patient recovery times, and pose life-threatening risks [1][5].

1.3 Resistance Mechanisms Within Biofilms

Biofilms confer antibiotic resistance through several mechanisms. The EPS matrix restricts the diffusion of antibiotics, acting as a physical and chemical shield [4][5]. Within this matrix, bacteria may express enzymes that degrade antibiotics before they reach their target [4][5]. Additionally, biofilms contain “persister cells”—dormant cells that tolerate antibiotics by temporarily halting metabolic activity [4][5]. Once antibiotic treatment ceases, these cells can repopulate the infection [4][5]. The combination of structural defense, enzyme production, and cellular dormancy makes biofilms exceptionally difficult to eliminate and a top priority for therapeutic innovation [4][5].

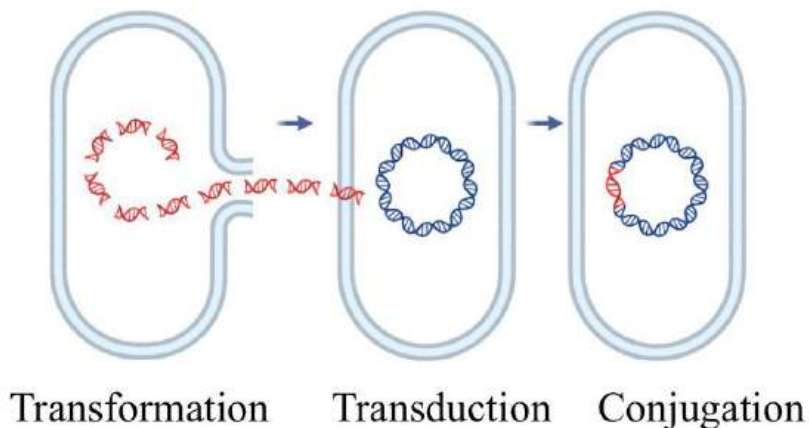
2. Horizontal Gene Transfer and Resistance Evolution

2.1 What Is Horizontal Gene Transfer?

Horizontal gene transfer (HGT) enables bacteria to acquire new genetic material—including antibiotic resistance genes—from other organisms or their environment [7–9]. This transfer occurs via three main pathways:

- Transformation: uptake of free DNA fragments from the environment [7].
- Transduction: transfer of DNA by bacteriophages (viruses that infect bacteria) [7].
- Conjugation: direct transfer of plasmids (small circular DNA molecules) between bacteria through cell-to-cell contact [8].

Unlike vertical transmission through reproduction, HGT allows for the immediate spread of resistance traits, even across species and genera [7–9]. This ability enables bacterial populations to rapidly adapt to antibiotic pressure, creating new multidrug-resistant strains seemingly overnight [2][3].



Structure #2 Process of horizontal gene transfer

Pathways of horizontal gene transfer (HGT). Bacteria exchange genetic material through transformation (uptake of free DNA), transduction (virus-mediated transfer), and conjugation (direct plasmid transfer). These mechanisms accelerate the spread of antibiotic resistance across species and environments.

2.2 Clinical Consequences of HGT

HGT plays a central role in the global spread of antibiotic resistance [7–9]. It is responsible for the emergence of so-called “superbugs,” such as CRE and MRSA, which are resistant to nearly all available antibiotics [2][3]. Hospitals have reported outbreaks where HGT has facilitated rapid transmission of resistance genes among patients, leading to quarantines and treatment failures [7][8]. These events underscore the urgency of developing diagnostic tools and containment strategies to identify and limit resistance transmission [1][2].

2.3 Resistance Gene Evolution and Multidrug Threats

HGT not only spreads resistance but also accelerates the evolution of complex, multidrug-resistant organisms [7–9]. Plasmids often carry multiple resistance genes, allowing bacteria to survive combination therapies that previously ensured effectiveness [8][9]. For example, resistance to colistin—a last-resort antibiotic—has emerged due to plasmid-mediated gene transfer [8]. These developments complicate diagnostics, limit therapeutic options, and increase the likelihood of treatment failure in both hospital and community settings [1–3].

3. Limited Development of New Antibiotics

3.1 Scientific and Structural Challenges

Developing new antibiotics is a scientifically rigorous and time-intensive process [10]. Bacterial cells have unique structural and biochemical features that make them difficult to target selectively without harming human cells [2][10]. Moreover, bacteria can rapidly develop resistance through mutation or HGT, rendering new drugs ineffective soon after their release [2][3]. Many promising compounds fail due to poor safety profiles or rapid clearance from the body [10].

3.2 Economic Disincentives and Industry Withdrawal

Antibiotics are typically prescribed in short courses, making them far less profitable than medications for chronic diseases [10]. Additionally, stewardship programs restrict the use of new antibiotics to preserve their effectiveness, further limiting financial returns [10]. These market conditions have led many pharmaceutical companies to abandon antibiotic development altogether [10]. Between 2000 and 2020, only a handful of new antibiotic classes have been introduced, leaving a dangerously thin pipeline [10].

3.3 Impact on Global Health

The slow pace of innovation has serious public health implications [1][10]. Procedures such as chemotherapy, organ transplantation, and joint replacement rely on effective antibiotics to prevent infection [1]. Without new treatments, these once-routine interventions could become too risky to perform [1][10]. The CDC estimates that antibiotic-resistant infections affect over 2.8 million people annually in the U.S., with more than 35,000 resulting in death [1]. This toll is projected to rise unless new interventions are developed and implemented globally [1][2].

4. Gene Editing to Reverse Resistance

4.1 What Is Gene Editing for Bacteria?

Gene editing technologies like CRISPR-Cas systems allow researchers to target and modify bacterial DNA with high precision [11–13]. In the context of antibiotic resistance, CRISPR can be engineered to recognize and cut resistance genes, especially those carried on plasmids or chromosomal DNA [11–13]. These systems can be delivered using bacteriophages, liposomes, or nanoparticles, offering the potential to make resistant bacteria susceptible to antibiotics once again [11–13].

4.2 Why Gene Editing Matters

Unlike traditional antibiotics, which often harm both harmful and beneficial bacteria, gene editing offers specificity [11–13]. CRISPR-based approaches can eliminate resistance genes while preserving the overall microbiome [11–13]. This strategy reduces the selective pressure that drives resistance and may prevent its emergence altogether [11–13]. Gene editing can also be tailored to match the genetic makeup of localized bacterial outbreaks, offering customized treatment solutions [12–13].

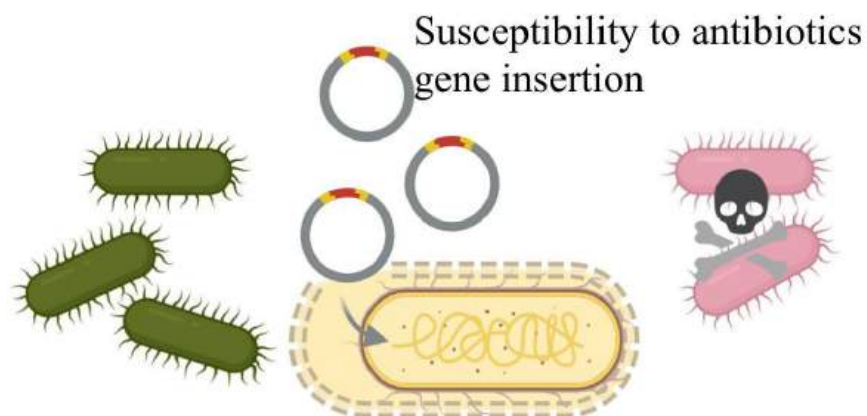


Figure #3: CRISPR-Cas process

CRISPR-Cas mechanism targeting resistance genes. By programming guide RNAs, CRISPR systems can selectively cut antibiotic resistance genes carried on plasmids or chromosomes, potentially restoring bacterial susceptibility to antibiotics.

4.3 Evidence and Specific Successes

A landmark study by Yosef et al. demonstrated that CRISPR-engineered bacteriophages targeting bla_{NDM-1} genes successfully eradicated multidrug-resistant bacteria [11–13]. In vivo experiments in mice showed that CRISPR-carrying phages reduced MRSA infections by over 90%, illustrating the potential of this approach in living systems [11–13]. CRISPR has also been used to sensitize resistant bacteria to antibiotics such as penicillin and carbapenems, restoring the efficacy of once-failed treatments [11–13].

4.4 Limitations

The greatest challenge in CRISPR therapy is effective delivery [11–13]. The immune system may neutralize delivery vectors, or they may fail to reach the infection site [11–13]. Infections that are widespread rather than localized are also more difficult to treat with CRISPR [12–13]. Additionally, identifying emerging resistance genes remains a time-consuming process, limiting the responsiveness of gene editing tools [12–13]. Despite these challenges, gene editing represents one of the most promising breakthroughs in the fight against resistance [11–13].

5. Quorum Sensing Inhibitors (QSIs): Blocking Bacterial Communication

5.1 What Is Quorum Sensing?

Quorum sensing is a communication mechanism bacteria use to coordinate group behavior based on population density [14–16]. They release signaling molecules, such as acyl homoserine lactones (AHLs) or autoinducing peptides (AIPs), which accumulate as cell numbers grow [14–16]. When a threshold concentration is reached, these molecules activate genes involved in biofilm formation, virulence factor production, and resistance expression [14–16]. QSIs are compounds that disrupt this signaling process, effectively silencing bacterial communication [17].

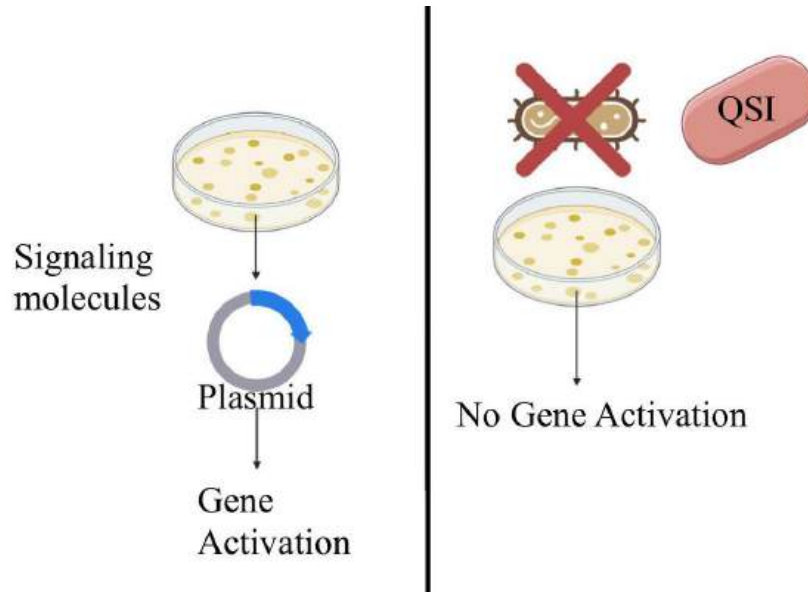


Figure #4: Quorum Sensing process with and without quorum sensing inhibitors

Bacterial quorum sensing. As bacterial populations grow, signaling molecules accumulate and trigger gene activation for biofilm formation and virulence. Quorum-sensing inhibitors (QSIs) block this communication, reducing bacterial defenses without directly killing cells.

5.2 Why QSIs Are Valuable

QSIs do not kill bacteria directly [17]. Instead, they weaken bacterial defenses, making them more susceptible to the immune system and conventional antibiotics [17]. By targeting non-essential processes like communication rather than survival, QSIs exert less selective pressure, reducing the risk of resistance development [17]. This makes them ideal candidates for combination therapies [15–16].

5.3 Evidence and Specific Applications

Natural QSIs, such as ajoene derived from garlic, have shown the ability to reduce biofilm density and downregulate virulence genes [15–17]. Laboratory studies confirm that QSIs are effective in both Gram-negative and Gram-positive bacteria, disrupting quorum sensing and reducing resistance gene expression [16–17]. When paired with antibiotics, these inhibitors significantly improve treatment outcomes [16–17].

5.4 Limitations

QSIs alone are often insufficient to eliminate infections [17]. Their efficacy depends on the pathogen and the complexity of its signaling networks [16]. For instance, *Pseudomonas aeruginosa* uses multiple quorum sensing systems, complicating inhibitor development [16]. Nonetheless, combining QSIs with antibiotics can greatly enhance therapeutic effectiveness, making this a key area for future development [16–17].

6. Artificial Intelligence: A New Frontier in Antibiotic Discovery

6.1 The Role of AI in Antibiotic Research

Artificial intelligence (AI) uses machine learning to analyze massive datasets, including genomic sequences, chemical compound libraries, and clinical records [18–19]. AI can detect patterns and correlations that humans may overlook, accelerating the discovery of new drug candidates, predicting resistance pathways, and optimizing existing treatment regimens [18–19].

6.2 Why AI Is Crucial

Traditional drug discovery processes take years and cost millions [18–19]. AI dramatically shortens this timeline by identifying promising molecules in a matter of weeks and predicting their efficacy against resistant strains [18–19]. AI also supports the development of personalized treatments by analyzing individual patient data and tailoring therapies accordingly [18–19].

6.3 Breakthroughs and Future Potential

In 2020, researchers used a deep learning algorithm to identify halicin, a novel antibiotic effective against resistant strains like *Acinetobacter baumannii* and MRSA [18–19]. Halicin's structure was unlike any existing antibiotic, demonstrating the power of AI to innovate beyond human intuition [18–19]. AI is also being used to model resistance trends and design proactive treatment strategies [18–19].

6.4 Limitations

Despite its promise, AI depends heavily on high-quality datasets [18–19]. Errors in input data can lead to inaccurate predictions [18–19]. Moreover, AI-generated drug candidates still require laboratory validation, a process that takes time and may allow resistance to continue evolving [18–19]. Still, AI is poised to transform antibiotic development by making it faster, more cost-effective, and more precise [18–19].

Conclusion

Antibiotic resistance remains one of the most pressing threats to global health, driven by the formation of protective biofilms, the rapid spread of resistance through horizontal gene transfer, and the dwindling pipeline of new antibiotics [1–5,7–10]. However, emerging innovations offer hope [11–19]. Gene editing holds the potential to precisely eliminate resistance genes [11–13]. Quorum sensing inhibitors can prevent bacterial communication and biofilm formation [14–17], while artificial intelligence is revolutionizing how we discover and optimize antibiotics [18–19]. While each solution has limitations, their combined power paves the way for a more targeted, efficient, and sustainable approach to managing bacterial infections [11–19]. The path forward demands continued research, investment, and interdisciplinary collaboration to preserve the efficacy of antibiotics for future generations [1–5,7–10].

Works Cited

- 1.Ventola, C. Lee. "The Antibiotic Resistance Crisis: Part 1: Causes and Threats." *P & T: A Peer-Reviewed Journal for Formulary Management*, vol. 40, no. 4, 2015, pp. 277-283.
<https://www.ncbi.nlm.nih.gov/pubmed/25859123>
- 2.Munita, Jose M., and Cesar A. Arias. "Mechanisms of Antibiotic Resistance." *Microbiology Spectrum*, vol. 4, no. 2, 2016, doi:10.1128/microbiolspec.VMBF-0016-2015.
<https://doi.org/10.1128/microbiolspec.VMBF-0016-2015>
- 3.Reygaert, Wanda C. "An Overview of the Antimicrobial Resistance Mechanisms of Bacteria." *AIMS Microbiology*, vol. 4, no. 3, 26 June 2018, pp. 482-501,
doi:10.3934/microbiol.2018.3.482. <https://doi.org/10.3934/microbiol.2018.3.482>
- 4.Rather, Muzamil Ahmad, et al. "Microbial Biofilm: Formation, Architecture, Antibiotic Resistance, and Control Strategies." *Brazilian Journal of Microbiology*, vol. 52, no. 4, 2021, pp. 1701-1718, doi:10.1007/s42770-021-00624-x.
<https://doi.org/10.1007/s42770-021-00624-x>
- 5.Roy, Ranita, et al. "Strategies for Combating Bacterial Biofilms: A Focus on Anti-Biofilm Agents and Their Mechanisms of Action." *Virulence*, vol. 9, no. 1, 2018, pp. 522-554,
doi:10.1080/21505594.2017.1313372. <https://doi.org/10.1080/21505594.2017.1313372>
- 6.Jeong, Hyeonsoo, et al. "Horizontal Gene Transfer in Human-Associated Microorganisms Inferred by Phylogenetic Reconstruction and Reconciliation." *Scientific Reports*, vol. 9, no. 1, 2019, article 5953, doi:10.1038/s41598-019-42227-5.
<https://www.nature.com/articles/s41598-019-42227-5>
- 7.Carattoli, Alessandra. "Plasmids and the Spread of Resistance." *International Journal of Medical Microbiology*, vol. 303, issues 6-7, 2013, pp. 298-304,
doi:10.1016/j.ijmm.2013.02.001. <https://doi.org/10.1016/j.ijmm.2013.02.001>
- 8.Kogay, Roman, et al. "Defence Systems and Horizontal Gene Transfer in Bacteria." *Environmental Microbiology*, vol. 26, no. 4, 2024, e16630,
doi:10.1111/1462-2920.16630.
<https://onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.16630>
- 9.Davies, Julian. "Where Have All the Antibiotics Gone?" *The Canadian Journal of Infectious Diseases & Medical Microbiology*, vol. 17, no. 5, 2006, pp. 287-290,
doi:10.1155/2006/707296. <https://doi.org/10.1155/2006/707296>
- 10.Allen, Alexander G., et al. "Gene Editing of HIV-1 Co-receptors to Prevent and/or Cure Virus Infection." *Frontiers in Microbiology*, vol. 9, article 2940, 17 Dec. 2018,
doi:10.3389/fmicb.2018.02940. <https://doi.org/10.3389/fmicb.2018.02940>
- 11.Aubert, Martine, et al. "Gene Editing for Latent Herpes Simplex Virus Infection Reduces Viral Load and Shedding In Vivo." *Nature Communications*, vol. 15, no. 1, article 4018, 13 May 2024, doi:10.1038/s41467-024-47940-y.
<https://doi.org/10.1038/s41467-024-47940-y>

12. Pursey, E., D. Sünderhauf, W. H. Gaze, E. R. Westra, and S. van Houte. "CRISPR-Cas Antimicrobials: Challenges and Future Prospects." *PLoS Pathogens*, vol. 14, no. 6, 2018, e1006990. <https://doi.org/10.1371/journal.ppat.1006990>
13. Mukherjee, Sampri, and Bonnie L. Bassler. "Bacterial Quorum Sensing in Complex and Dynamically Changing Environments." *Nature Reviews Microbiology*, vol. 17, no. 6, 2019, pp. 371-382, doi:10.1038/s41579-019-0186-5. <https://doi.org/10.1038/s41579-019-0186-5>
14. Azimi, Sheyda, et al. "Bacterial Quorum Sensing During Infection." *Annual Review of Microbiology*, vol. 74, 2020, pp. 201-219, doi:10.1146/annurev-micro-032020-093845. <https://doi.org/10.1146/annurev-micro-032020-093845>
15. Tiwari, Sakshi, et al. "Quorum Sensing in Gram-Negative Bacteria: Strategies to Overcome Antibiotic Resistance in Ocular Infections." *Current Molecular Medicine*, vol. 24, no. 7, 2024, pp. 876-888, doi:10.2174/1566524023666230727094635. <https://doi.org/10.2174/1566524023666230727094635>
16. Kalia, Vipin Chandra. "Quorum Sensing Inhibitors: An Overview." *Biotechnology Advances*, vol. 31, no. 2, 2013, pp. 224-245, doi:10.1016/j.biotechadv.2012.10.004. <https://doi.org/10.1016/j.biotechadv.2012.10.004>
17. Gangwal, Amit, and Antonio Lavecchia. "Unleashing the Power of Generative AI in Drug Discovery." *Drug Discovery Today*, vol. 29, no. 6, 2024, article 103992, doi:10.1016/j.drudis.2024.103992. <https://doi.org/10.1016/j.drudis.2024.103992>
18. Bhowmick, Mithun, et al. "Future Prospective of AI in Drug Discovery." *Advances in Pharmacology*, no. 103, 2025, pp. 429-449, doi:10.1016/bs.apha.2025.01.009. <https://doi.org/10.1016/bs.apha.2025.01.009>
19. Paul, D., G. Sanap, S. Shenoy, S. Kalyane, K. Kalia, R. K. Tekade. "Artificial Intelligence in Drug Discovery and Development." *Drug Discovery Today*, vol. 26, no. 1, 2021, pp. 80-93, doi:10.1016/j.drudis.2020.10.010. <https://doi.org/10.1016/j.drudis.2020.10.010>

Brain Regions and Neural Networks that Encode Semantic Memory By Milton Dai

Abstract

Memory is defined as an essential cognitive function that permits individuals to acquire, retain, and recover data (Sridhar et al.). There are three types of memory: working memory, non-declarative memory, and declarative memory. Declarative memory is associated with encoding facts, knowledge, and personal experiences. It is composed of semantic (facts) and episodic (experiences) memory. While region and network mapping are well established for memory in general, a consensus about the brain regions and neural networks that encode semantic memory has yet to be established. Thus, this paper investigates the brain regions that encode semantic memory, and how they relate with episodic memory. Using two fMRI data repositories, NeuroQuery and NeuroSynth, activation and differential maps were generated for these two types of memory. Subsequently, the data set was parsed for regions which are associated with only semantic memory.

Introduction

Cognitive neuroscience broadly defines memory as the capacity to store and retrieve information (Zlotnik and Vansintjan). Furthermore, memory is often divided into two major types: working memory (short-term memory) and long-term memory. Long-term memory can be further categorized into declarative memory and non-declarative memory. Nondeclarative memory involves the unconscious retention of information, and is expressed through performance rather than recollection (Squire et al.). Conversely, declarative memory refers to knowledge about facts and events (Sridhar et al.) and will be the primary focus of this study.

There are two types of declarative memory: episodic and semantic memory. Episodic memory is recollection of one's previous experiences. Episodic memory has been relatively well characterized, and the brain regions associated with encoding episodic memory include the hippocampus, posterior cingulate cortex, and the ventral parietal cortex (Rugg and Vilberg). Semantic memory involves the retention of facts. There is currently no consensus of where semantic memory is encoded in the brain, but it has been theorized that perceptual and motor systems play a predominant role alongside the hippocampus in encoding this type of memory (Sridhar et al.).

Several theories have been proposed on the brain systems that store semantic memory. Some suggest that semantic and episodic memory use similar neural networks located in the medial temporal lobes and hippocampal regions (Vargha-Khadem et al.), but others suggest a more universal activation pattern across the cerebral cortex (Zhang et al.). Moreover, Frisby et al. summarize three main frameworks on how semantic information is encoded. One of the predominant theories of semantic memory is called "convergence zones", first proposed by Damasio, which suggested that semantic activation depended on neural activity in multiple regions simultaneously (Gainotti). Another model is a modality-distributed semantics model in which modality-specific brain regions encode and store semantic memory (Tong et al.). The final

framework is a “hub-and-spoke” hypothesis proposed by Patterson et al. which adds a central “hub”, thought to be located in the anterior temporal lobes, responsible for interpreting the modality specific “spokes” from the modality-distributed model. Lambon Ralph et al. recently made adjustments to the aforementioned “hub-and-spoke” model, adding an element of “semantic control” in addition to “semantic representation” (hub-and-spokes), implemented by a different neural network centered around the prefrontal cortex (Chiou et al.). Thus, it can be argued that the controlled semantic cognition model (CSC) is a hybrid theory that combines elements of both the hub-and-spoke hypothesis and the modality-distributed semantics model.

A lack of consensus on brain regions and neural networks that encode semantic memory in the brain has led to the question: Which brain regions and networks map to semantic memory? Functional magnetic resonance imaging (fMRI) is a neuroimaging technique based on MRI that uses nuclear magnetic resonance to measure changes in brain metabolism as blood flow shifts to activated regions (Whitten; Glover). Using NeuroQuery and NeuroSynth, two neuroimaging meta-analysis tools that generate brain maps based on studies performed with fMRI, a comparison was performed between the regions associated with general, semantic and episodic memory. Consequently, an area that best demonstrated the various neural networks associated with semantic memory was determined.

Methods

Data Sources

Data was collected using the meta-analysis tools NeuroQuery (Duckett et al.) and NeuroSynth (Yarkoni et al.). These contain large repositories of fMRI data which are used to generate brain activation maps based on the current research environment. Both tools were used in this study to ensure greater reliability of the findings.

Data Collection

For each tool, brain maps for semantic memory, episodic memory, and declarative memory were generated. These were then downloaded and loaded into Python. To generate differential maps, maps were aligned spatially using the *resample_to_img* function from package NiLearn and thresholded to keep only activations above 2.0 using the *math_img* function. Difference maps were then generated using the *math_img* function. The code can be found in the Appendix.

Statistics

Z-statistics were converted into two-tailed p-values with *scipy.stats.norm.sf*. Only maps from NeuroQuery were thresholded using FDR correction ($q < 0.01$), as NeuroSynth maps were already FDR-corrected at $q < 0.01$. Comparing significant voxels with the Harvard-Oxford atlas allowed regions and significant voxels to be labelled.

Results

To explore differences between semantic and episodic memory, brain activation maps and difference maps were compared between the two types of declarative memory, leading to several key findings. Figures 1a-d show brain activation maps for the term “semantic memory” in both NeuroSynth (Fig 1a) and NeuroQuery (Fig 1b) databases. Both figures display Blood-Oxygen-Level-Dependent (BOLD) fMRI activation in the left temporal lobe as that was the region that showed the most activation, but Figure 1b and 1d center around the hippocampus and the medial temporal lobes (more medial coordinate activation) as that is a region heavily implicated in episodic memory.

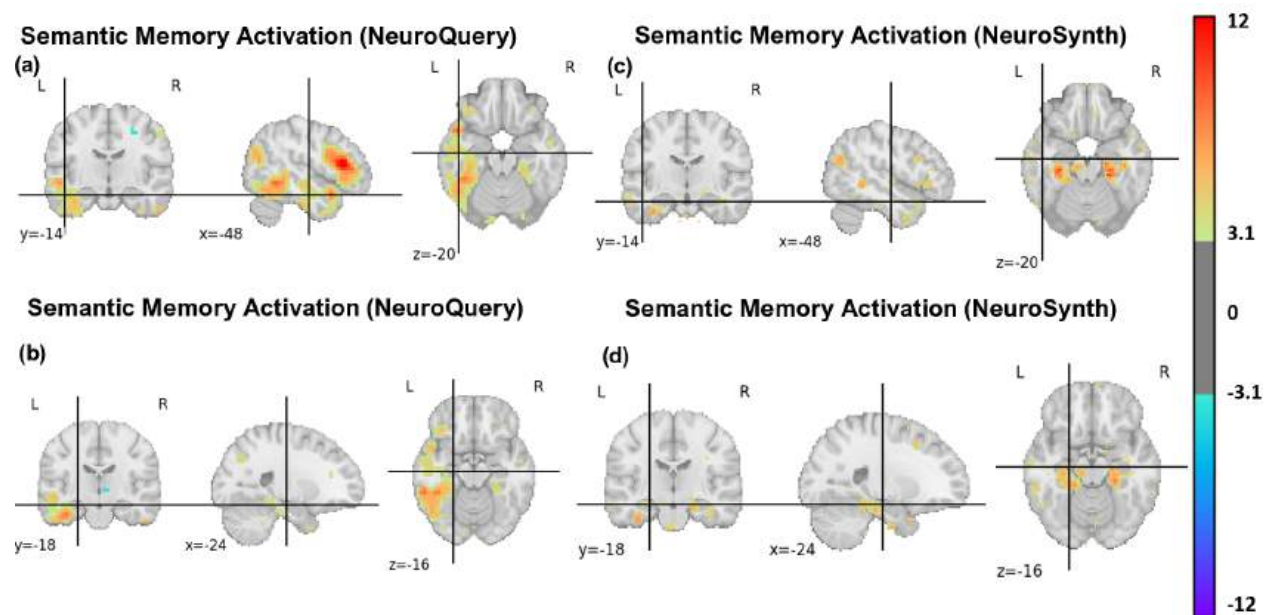


Figure 1a-d: Brain activation maps for term “semantic memory” from NeuroSynth and NeuroQuery from coordinates comparing anterior and posterior regions around the hippocampal areas. Figures display coronal, sagittal, transverse slices (left to right); color bar shows z-statistic. 1a-b: Common activations are seen in the occipital lobe, prefrontal cortex, and across the temporal lobe. Z-statistics (Figure 1a-b): min 3.64, max 11.11; P-value $p < 0.05$ FDR corrected $q < 0.01$. Z-statistics (Figure 1c-d): min -3.69, max 13.91; P-value $p < 0.05$ FDR corrected $q < 0.05$. For visualization purposes, only voxels exceeding a z-statistic of 3.1 and below -3.1 were displayed.

NeuroSynth (Fig. 1a-b)		NeuroQuery (Fig. 1c-d)	
Brain Area	Number of Active Voxels ($p < 0.05$)	Brain Area	Number of Active Voxels ($p < 0.05$)
Frontal Pole	354 voxels	Frontal Pole	11 voxels
Insular Cortex	20 voxels	Superior Frontal Gyrus	7 voxels
Superior Frontal Gyrus	34 voxels	Middle Frontal Gyrus	10 voxels
Middle Frontal Gyrus	92 voxels	Inferior Frontal Gyrus, pars triangularis	69 voxels

Inferior Frontal Gyrus, pars triangularis	68 voxels	Inferior Frontal Gyrus, pars opercularis	71 voxels
Inferior Frontal Gyrus, pars opercularis	58 voxels	Precentral Gyrus	39 voxels
Temporal Pole	630 voxels	Temporal Pole	59 voxels
Superior Temporal Gyrus, anterior division	9 voxels	Superior Temporal Gyrus, anterior division	4 voxels
Superior Temporal Gyrus, posterior division	3 voxels	Superior Temporal Gyrus, posterior division	25 voxels
Middle Temporal Gyrus, anterior division	109 voxels	Middle Temporal Gyrus, anterior division	14 voxels
Middle Temporal Gyrus, posterior division	182 voxels	Middle Temporal Gyrus, posterior division	56 voxels
Middle Temporal Gyrus, temporooccipital part	118 voxels	Middle Temporal Gyrus, temporooccipital part	55 voxels
Inferior Temporal Gyrus, anterior division	104 voxels	Inferior Temporal Gyrus, anterior division	13 voxels
Inferior Temporal Gyrus, posterior division	84 voxels	Inferior Temporal Gyrus, posterior division	52 voxels
Inferior Temporal Gyrus, temporooccipital part	76 voxels	Inferior Temporal Gyrus, temporooccipital part	66 voxels
Postcentral Gyrus	10 voxels	Postcentral Gyrus	3 voxels
Supramarginal Gyrus, posterior division	3 voxels	Superior Parietal Lobule	7 voxels
Angular Gyrus	43 voxels	Supramarginal Gyrus, anterior division	1 voxels
Lateral Occipital Cortex, superior division	360 voxels	Angular Gyrus	3 voxels
Lateral Occipital Cortex, inferior division	13 voxels	Lateral Occipital Cortex, superior division	80 voxels
Intracalcarine Cortex	4 voxels	Lateral Occipital Cortex, inferior division	24 voxels
Frontal Medial Cortex	19 voxels	Intracalcarine Cortex	1 voxels
Subcallosal Cortex	15 voxels	Frontal Medial Cortex	1 voxels
Paracingulate Gyrus	15 voxels	Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)	2 voxels
Cingulate Gyrus, anterior division	2 voxels	Paracingulate Gyrus	20 voxels
Cingulate Gyrus, posterior division	128 voxels	Cingulate Gyrus, posterior division	2 voxels
Precuneous Cortex	466 voxels	Precuneous Cortex	7 voxels
Frontal Orbital Cortex	36 voxels	Frontal Orbital Cortex	42 voxels
Parahippocampal Gyrus, anterior division	70 voxels	Parahippocampal Gyrus, anterior division	3 voxels
Parahippocampal Gyrus, posterior division	448 voxels	Parahippocampal Gyrus, posterior division	16 voxels

Lingual Gyrus	67 voxels	Temporal Fusiform Cortex, anterior division	11 voxels
Temporal Fusiform Cortex, anterior division	155 voxels	Temporal Fusiform Cortex, posterior division	94 voxels
Temporal Fusiform Cortex, posterior division	477 voxels	Temporal Occipital Fusiform Cortex	23 voxels
Temporal Occipital Fusiform Cortex	9 voxels	Occipital Fusiform Gyrus	7 voxels
Frontal Opercular Cortex	22 voxels	Frontal Opercular Cortex	6 voxels
Central Opercular Cortex	7 voxels	Frontal Pole	11 voxels
Parietal Opercular Cortex	1 voxels	Superior Frontal Gyrus	7 voxels
Planum Polare	11 voxels	Middle Frontal Gyrus	10 voxels
Heschl's Gyrus (includes H1 and H2)	3 voxels		
Planum Temporale	5 voxels		
Occipital Pole	4 voxels		

Table 1: Brain regions showing significant voxel activity in NeuroSynth and NeuroQuery brains for term “semantic memory”. Note that significant voxel activity indicates the existence of activity, not necessarily the strength of activity.

The most salient activation areas in the NeuroQuery map were in the occipital lobe, fusiform gyrus, inferior frontal lobe, and in the left temporal lobe (Fig. 1a $p\text{-val} < 0.05$). Despite the large number of active voxels as opposed to the NeuroQuery map, the NeuroSynth map shows similar, weaker, activations across matching regions, while displaying stronger activation in the hippocampus (Fig 1a-b $p\text{-val} < 0.05$; Table 1). These distinctions mirror the method of data collection of both databases: NeuroSynth uses text-mining to parse existing data and thus shows strong levels of activation in brain regions that have received a profusion of research (i.e. hippocampus), while NeuroQuery is a predictive model and thus shows more even activations across the whole brain. Still, the large amount of brain regions activated reflect a wide range of activation for semantic memory. It is also vital to note the small, disparate areas of activation in the NeuroSynth map (Figure 1c $p\text{-val} < 0.05$, Table 1) ranging from the frontal lobe to the occipital lobe. While this supports the notion that a multitude of brain regions are involved in encoding and storing semantic memory, it also suggests that a widely distributed neural network not limited to specific brain regions are implicated in semantic memory.

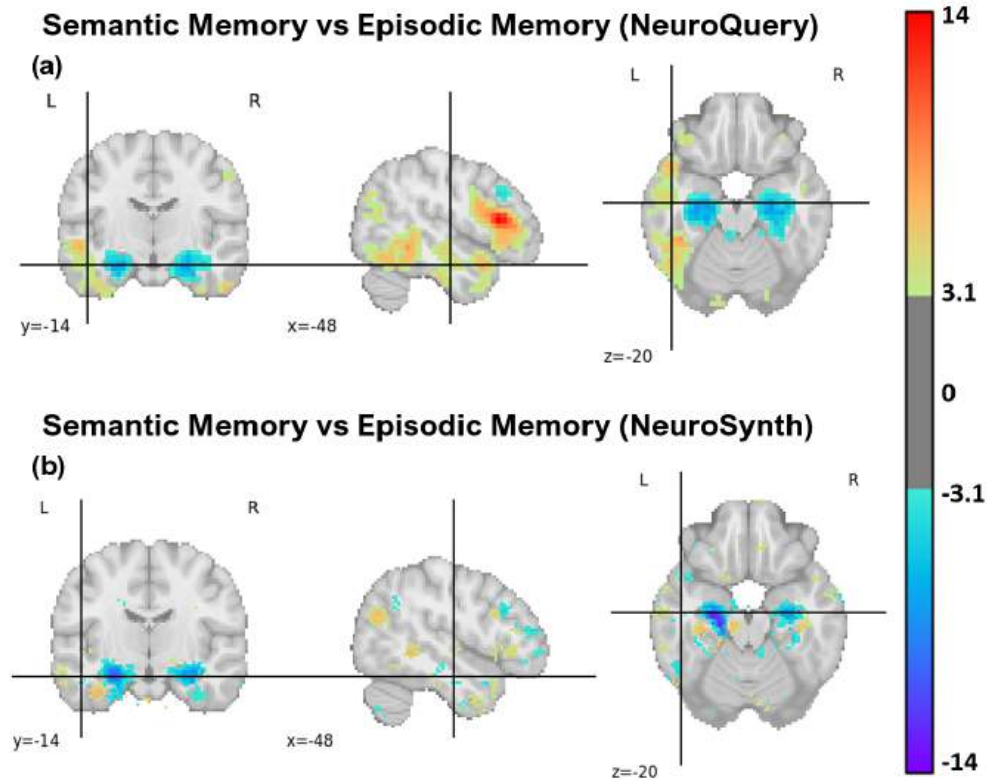


Figure 2: Difference map between terms “semantic memory” and “episodic memory” Z-maps from NeuroSynth and NeuroQuery. Figures display coronal, sagittal, transverse slices (left to right); color bar shows z-statistic. Z-statistics (Figure 1a) min: -9.45, max: 13.91; P-value $p < 0.05$ FDR corrected $q < 0.01$. Z-statistics (Figure 1b): min -13.27, max 8.96; P-value $p < 0.05$ FDR corrected $q < 0.01$. Consistent differences in the medial temporal lobes (shown by blue-purple colors) and lateral temporal lobe (shown by green-red colors) indicate some difference in the terms’ neural substrates. For visualization purposes, only voxels exceeding a z-statistic of 3.1 and below -3.1 were displayed.

The difference maps between episodic memory and semantic memory revealed further patterns. Both NeuroSynth and NeuroQuery maps displayed a recurrent resemblance: heavy episodic activation in the medial temporal lobes and hippocampus (Fig 2a-b $p\text{-val} < 0.05$; Table 2). Despite activation in these areas in the maps for semantic memory, it is clear that episodic memory supersedes its semantic counterpart, implying that while these brain regions are involved in both forms of memory, it is more strongly associated with episodic activation. Coupled with the strong activations across the left temporal lobe in the NeuroQuery map (Fig 2a $p\text{-val} < 0.05$; Table 2), this may reflect a wide range of neural networks that encode and store semantic memory, rather than activation in localized regions.

Comparisons between general memory and semantic memory further suggest that semantic memory may be encoded across a multitude of brain regions. The maps from both databases exhibit numerous differences, likely in part due to discrepancies between the databases for NeuroSynth and NeuroQuery (Fig 3a-b $p\text{-val} < 0.05$; Table 3). The NeuroQuery map displays strong semantic activation across the left lateral temporal lobe, inferior frontal cortex, and

fusiform gyrus and little memory activation (Fig 3a p-val < 0.05; Table 3). Meanwhile, the NeuroSynth map shows activation for memory in the hippocampus, dorsal region of the frontal cortex, and dispersed areas of activation for semantic memory (Fig 3b p-val < 0.05; Table 3).

NeuroSynth		NeuroQuery	
Regions	Number of Active Voxels	Regions	Number of Active Voxels
Frontal Pole	879 voxels	Frontal Pole	137 voxels
Insular Cortex	20 voxels	Insular Cortex	12 voxels
Superior Frontal Gyrus	170 voxels	Superior Frontal Gyrus	75 voxels
Middle Frontal Gyrus	383 voxels	Middle Frontal Gyrus	106 voxels
Inferior Frontal Gyrus, pars triangularis	98 voxels	Inferior Frontal Gyrus, pars triangularis	73 voxels
Inferior Frontal Gyrus, pars opercularis	79 voxels	Inferior Frontal Gyrus, pars opercularis	80 voxels
Precentral Gyrus	7 voxels	Precentral Gyrus	90 voxels
Temporal Pole	687 voxels	Temporal Pole	163 voxels
Superior Temporal Gyrus, anterior division	9 voxels	Superior Temporal Gyrus, anterior division	25 voxels
Superior Temporal Gyrus, posterior division	3 voxels	Superior Temporal Gyrus, posterior division	46 voxels
Middle Temporal Gyrus, anterior division	132 voxels	Middle Temporal Gyrus, anterior division	57 voxels
Middle Temporal Gyrus, posterior division	231 voxels	Middle Temporal Gyrus, posterior division	111 voxels
Middle Temporal Gyrus, temporooccipital part	116 voxels	Middle Temporal Gyrus, temporooccipital part	87 voxels
Inferior Temporal Gyrus, anterior division	117 voxels	Inferior Temporal Gyrus, anterior division	25 voxels
Inferior Temporal Gyrus, posterior division	99 voxels	Inferior Temporal Gyrus, posterior division	103 voxels
Inferior Temporal Gyrus, temporooccipital part	128 voxels	Inferior Temporal Gyrus, temporooccipital part	77 voxels
Postcentral Gyrus	16 voxels	Postcentral Gyrus	13 voxels
Superior Parietal Lobule	2 voxels	Superior Parietal Lobule	25 voxels
Supramarginal Gyrus, posterior division	6 voxels	Supramarginal Gyrus, anterior division	14 voxels
Angular Gyrus	140 voxels	Supramarginal Gyrus, posterior division	22 voxels
Lateral Occipital Cortex, superior division	1021 voxels	Angular Gyrus	63 voxels

Lateral Occipital Cortex, inferior division	17 voxels	Lateral Occipital Cortex, superior division	364 voxels
Intracalcarine Cortex	11 voxels	Lateral Occipital Cortex, inferior division	90 voxels
Frontal Medial Cortex	22 voxels	Intracalcarine Cortex	7 voxels
Subcallosal Cortex	15 voxels	Frontal Medial Cortex	1 voxels
Paracingulate Gyrus	92 voxels	Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)	12 voxels
Cingulate Gyrus, anterior division	13 voxels	Paracingulate Gyrus	39 voxels
Cingulate Gyrus, posterior division	471 voxels	Cingulate Gyrus, anterior division	1 voxels
Precuneous Cortex	1449 voxels	Cingulate Gyrus, posterior division	61 voxels
Cuneal Cortex	33 voxels	Precuneous Cortex	381 voxels
Frontal Orbital Cortex	121 voxels	Cuneal Cortex	75 voxels
Parahippocampal Gyrus, anterior division	355 voxels	Frontal Orbital Cortex	83 voxels
Parahippocampal Gyrus, posterior division	385 voxels	Parahippocampal Gyrus, anterior division	85 voxels
Lingual Gyrus	273 voxels	Parahippocampal Gyrus, posterior division	80 voxels
Temporal Fusiform Cortex, anterior division	191 voxels	Lingual Gyrus	58 voxels
Temporal Fusiform Cortex, posterior division	560 voxels	Temporal Fusiform Cortex, anterior division	20 voxels
Temporal Occipital Fusiform Cortex	104 voxels	Temporal Fusiform Cortex, posterior division	132 voxels
Occipital Fusiform Gyrus	2 voxels	Temporal Occipital Fusiform Cortex	55 voxels
Frontal Opercular Cortex	22 voxels	Occipital Fusiform Gyrus	24 voxels
Central Opercular Cortex	7 voxels	Frontal Opercular Cortex	14 voxels
Parietal Opercular Cortex	1 voxels	Parietal Opercular Cortex	2 voxels
		Planum Polare	3 voxels
		Supracalcarine Cortex	5 voxels
		Occipital Pole	16 voxels

Table 2: Brain regions showing significant voxel activity in NeuroSynth and NeuroQuery difference maps for terms “semantic memory” and “episodic memory”. Note that significant voxel activity indicates the existence of activity, not necessarily the strength of activity.

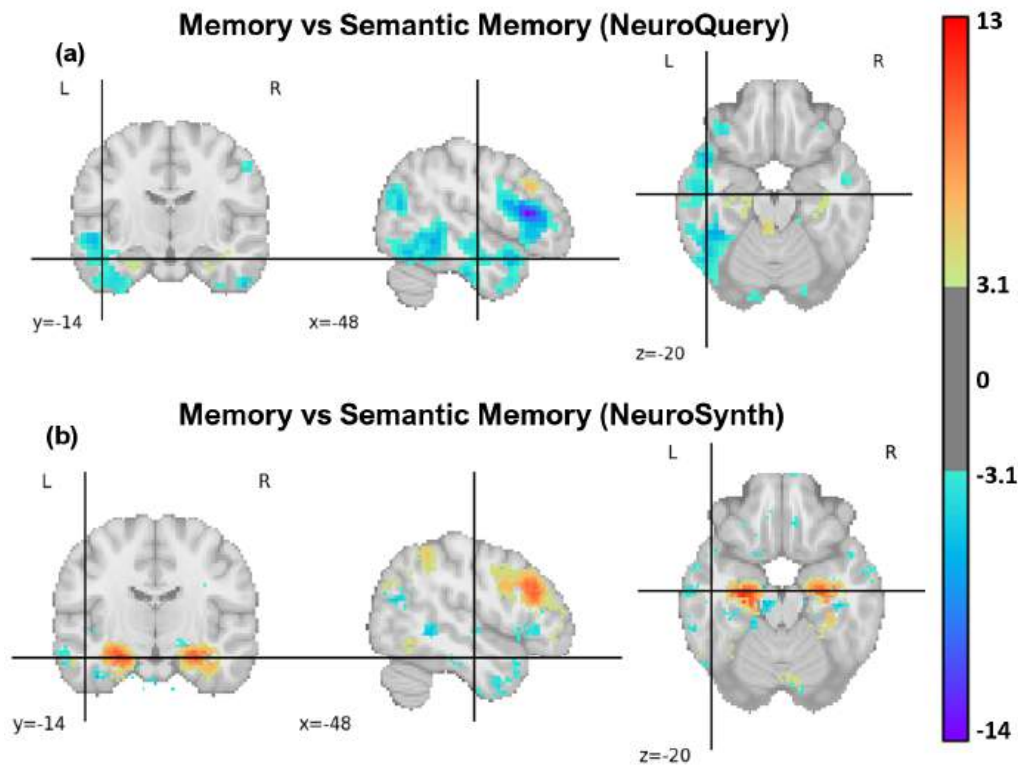


Figure 3: Difference map between terms “memory” and “semantic memory” z-maps from NeuroSynth and NeuroQuery. Z-statistics (Figure 1a): min: -13.91, max: 6.59; P-value $p < 0.05$ FDR corrected $q < 0.01$. Z-statistics (Figure 1b): min: -8.96, max: 12.35; P-value $p < 0.05$ FDR corrected $q < 0.01$. Strong activation for memory (red) in hippocampal areas demonstrates a weaker role for the hippocampus in storing semantic memory. For visualization purposes, only voxels exceeding a z-statistic of 3.1 and below -3.1 were displayed.

NeuroSynth		NeuroQuery	
Regions	Number of Active Voxels	Regions	Number of Active Voxels
Frontal Pole	1725 voxels	Frontal Pole	185 voxels
Insular Cortex	130 voxels	Insular Cortex	8 voxels
Superior Frontal Gyrus	570 voxels	Superior Frontal Gyrus	68 voxels
Middle Frontal Gyrus	2276 voxels	Middle Frontal Gyrus	128 voxels
Inferior Frontal Gyrus, pars triangularis	250 voxels	Inferior Frontal Gyrus, pars triangularis	76 voxels
Inferior Frontal Gyrus, pars opercularis	193 voxels	Inferior Frontal Gyrus, pars opercularis	79 voxels
Precentral Gyrus	219 voxels	Precentral Gyrus	90 voxels
Temporal Pole	663 voxels	Temporal Pole	162 voxels
Superior Temporal Gyrus, anterior division	9 voxels	Superior Temporal Gyrus, anterior division	25 voxels

Superior Temporal Gyrus, posterior division	3 voxels	Superior Temporal Gyrus, posterior division	46 voxels
Middle Temporal Gyrus, anterior division	136 voxels	Middle Temporal Gyrus, anterior division	61 voxels
Middle Temporal Gyrus, posterior division	195 voxels	Middle Temporal Gyrus, posterior division	115 voxels
Middle Temporal Gyrus, temporooccipital part	120 voxels	Middle Temporal Gyrus, temporooccipital part	86 voxels
Inferior Temporal Gyrus, anterior division	120 voxels	Inferior Temporal Gyrus, anterior division	23 voxels
Inferior Temporal Gyrus, posterior division	89 voxels	Inferior Temporal Gyrus, posterior division	106 voxels
Inferior Temporal Gyrus, temporooccipital part	186 voxels	Inferior Temporal Gyrus, temporooccipital part	78 voxels
Postcentral Gyrus	10 voxels	Postcentral Gyrus	14 voxels
Superior Parietal Lobule	295 voxels	Superior Parietal Lobule	22 voxels
Supramarginal Gyrus, anterior division	2 voxels	Supramarginal Gyrus, anterior division	14 voxels
Supramarginal Gyrus, posterior division	366 voxels	Supramarginal Gyrus, posterior division	22 voxels
Angular Gyrus	352 voxels	Angular Gyrus	47 voxels
Lateral Occipital Cortex, superior division	2236 voxels	Lateral Occipital Cortex, superior division	265 voxels
Lateral Occipital Cortex, inferior division	13 voxels	Lateral Occipital Cortex, inferior division	91 voxels
Intracalcarine Cortex	25 voxels	Intracalcarine Cortex	5 voxels
Frontal Medial Cortex	21 voxels	Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)	10 voxels
Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)	33 voxels	Paracingulate Gyrus	31 voxels
Subcallosal Cortex	15 voxels	Cingulate Gyrus, anterior division	1 voxels
Paracingulate Gyrus	381 voxels	Cingulate Gyrus, posterior division	23 voxels
Cingulate Gyrus, anterior division	2 voxels	Precuneous Cortex	104 voxels
Cingulate Gyrus, posterior division	638 voxels	Cuneal Cortex	37 voxels
Precuneous Cortex	1159 voxels	Frontal Orbital Cortex	79 voxels
Cuneal Cortex	19 voxels	Parahippocampal Gyrus, anterior division	27 voxels
Frontal Orbital Cortex	144 voxels	Parahippocampal Gyrus, posterior division	35 voxels

Parahippocampal Gyrus, anterior division	386 voxels	Lingual Gyrus	33 voxels
Parahippocampal Gyrus, posterior division	370 voxels	Temporal Fusiform Cortex, anterior division	25 voxels
Lingual Gyrus	300 voxels	Temporal Fusiform Cortex, posterior division	106 voxels
Temporal Fusiform Cortex, anterior division	166 voxels	Temporal Occipital Fusiform Cortex	50 voxels
Temporal Fusiform Cortex, posterior division	521 voxels	Occipital Fusiform Gyrus	24 voxels
Temporal Occipital Fusiform Cortex	134 voxels	Frontal Opercular Cortex	14 voxels
Occipital Fusiform Gyrus	16 voxels	Central Opercular Cortex	2 voxels
Frontal Opercular Cortex	24 voxels	Parietal Opercular Cortex	3 voxels
		Planum Polare	3 voxels
		Supracalcarine Cortex	1 voxels
		Occipital Pole	14 voxels

Table 3: Brain regions showing significant voxel activity in NeuroSynth and NeuroQuery difference maps for terms “memory” and “semantic memory”. Note that significant voxel activity indicates the existence of activity, not necessarily the strength of activity.

Discussion

3.1 Key Findings

This research revealed several key findings regarding the process of encoding and storing semantic memory in the brain. First, it was shown that a large number of brain regions are involved in encoding semantic memory, being against proposed theories that semantic and episodic memory are encoded using the same neural networks (Ryan et al.). These regions include the occipital, frontal, parietal, and temporal lobes, as well as the hippocampal areas, suggesting a distributed network of activation (Thompson-Schill). However, maps also provide evidence against the use of modality-specific systems in semantic memory, where different brain regions encode semantic memory depending on the type of sensory information it is associated with, as suggested by prior research (Binder and Desai). To illustrate, Figure 1 shows no significant activation for either of Heschl’s Gyrus, Primary Visual Cortex, or Somatosensory Cortex, which are associated with audition (Warrier et al.), vision, and somatosensation, respectively. Still, the activation of some modality-specific areas (inferior frontal gyrus, lingual gyrus, and posterior temporal cortex) provide some evidence that there is some multimodal involvement. Thus, these findings provide support for a controlled semantic cognition (CSC) model of semantic memory (Lambon Ralph et al.), a framework combining the Hub-and-Spoke model of semantic memory (Patterson et al.) with semantic control systems monitored by the prefrontal cortex. As referenced in the Introduction, in this model, semantic processing occurs in a “hub” in the semantic network, which is theorized to be located in the anterior temporal lobe (Patterson and Lambon Ralph). The semantic control system is responsible for modulating this

system depending on the task (Chiou et al.) . With Figure 1 showing activation across the temporal lobe, in regions including the inferior frontal gyrus, left middle temporal gyrus, and strong activation at the left anterior temporal lobe, this indicates the existence of a “hub” which integrates various inputs from modality-specific “spokes”. The strong activation in the prefrontal cortex also supports the existence of a semantic control system. This is further buttressed by the strong left-lateralized activation shown by the NeuroQuery Maps in Figure 1, a region strongly associated with language (Corballis). As language is key to integrating conceptual information (Reilly and Peelle), it is fitting that this region provides the hub for encoding semantic memory.

Additionally, the difference map between semantic and episodic memory indicated a difference in their neural substrates. Figure 2a-b shows strong patterns of activation for episodic memory around the medial temporal lobe and hippocampal regions, despite the presence of semantic activation in Figure 1. This implies that the role of these regions in encoding semantic memory, while present, is quite restricted, especially when compared with episodic memory. These findings reflect views that while the hippocampus and medial temporal lobes are involved in encoding episodic memory, they have a time-limited role in the formation of semantic knowledge (Manns et al.; Wixted et al.; Squire et al.). Moreover, creating the difference map had little effect on the left inferior frontal gyrus and left temporal gyrus, underscoring the relationship between semantic memory and language.

The difference map between semantic memory and memory also reveals distinctions between the activation patterns of general and semantic memory. Similar to the episodic-semantic map (Figure 2b), the NeuroSynth map (Figure 3b) displayed strong patterns of activation for general memory around the medial temporal lobes and hippocampal regions, reiterating the notion that semantic memory is a distributed process that does not solely rely on the hippocampus and areas encoding episodic memory. The NeuroQuery map (Figure 3a) also highlighted this fact, although to a lesser extent. Crucially, the map shows widespread regions of activation through the frontal and temporal lobes, emphasizing semantic memory’s different activation network compared to other forms of memory such as working or procedural memory, which are theorized to be stored in the same regions that store sensory information (Eriksson et al.), and the basal ganglia and cerebellum (Bouyeure and Noulhiane), respectively. However, a contradiction between the maps was found in the inferior frontal gyrus, in which the NeuroSynth map shows general memory activation, but the NeuroQuery map shows semantic activation. This may reflect the role of the left inferior frontal areas in other processes such as auditory comprehension and other working memory processes (Hickok and Poeppel). The function which contains more emphasis is then determined by the meta-analytic tool used to generate the map. Still, this heterogeneity points to a limitation of using meta-analytic tools such as NeuroSynth or NeuroQuery: they can only generate maps based on prior research, and the methodological approaches used by either tool is different as well. This will be discussed further in the Limitations section.

Impact

Understanding the neural substrates behind semantic memory has numerous real-world applications, from clinical interventions in medical conditions such as semantic aphasia, a language disorder that affects an individual's ability to communicate, to applications in education, such as developing more effective study techniques. Overall, the above maps provide evidence that semantic memory uses its own network of activation, different from general memory, but with some overlaps with episodic memory. Judging from activation and difference maps of semantic memory and other forms of memory, it was determined that the brain regions activated were most closely aligned to a semantic cognition model of memory. Specifically, a high level of activation in the left anterior temporal lobe showed evidence for a 'hub' in that region, with modality specific areas such as the lingual gyrus, fusiform gyrus, and posterior temporal cortex being potential locations for "spokes". Moreover, strong activation in the prefrontal cortex correlated with the existence of a semantic control system which monitors the Hub-and-Spoke system (Chiou et al.). The data also contradicted other models like the convergence-zones model, which advocated for increased right hemisphere activation (Gainotti), a pattern not shown in the activation maps. Despite the correlations with the semantic cognition model of semantic memory, there were some differences. Specifically, not all expected modality-specific "spokes" were activated in the brain maps. As mentioned previously, key areas for sensation such as Heschl's Gyrus, Primary Visual Cortex, or Somatosensory Cortex showed negligible activation, even though these are considered as parts of the "spokes" in the model. However, as spoke recruitment likely depends on task demands, this may simply reflect a paucity of research in those specific areas. Therefore, further research could focus on mapping these disparate regions in order to truly verify the semantic cognition model.

Limitations

There were several limitations during the composition of this research paper. Firstly, all activation maps were generated with BOLD fMRI, as opposed to other types such as ASL or CBV-weighted, which may have provided different advantages, such as higher temporal resolution or greater detail (Huber et al.; Borogovac and Asllani). Furthermore, despite the many advantages of fMRI, some weaknesses of this methodology should be discussed. The main limitation of fMRI lies in the fact that it can only reflect mass activity in the brain, thus it cannot detect underlying changes at the neuronal level (Logothetis). In other words, correlations of brain activity with a particular behaviour does not necessarily mean it is involved in that function (Whitten).

Crucially, there are additional constraints when using meta-analytic tools such as NeuroSynth or NeuroQuery. Firstly, the accuracy tools of these tools rely on the depth of the databases which are used to generate the activation maps, so the areas of activation may reflect current research trends instead of true activation and may contain publication bias. This is exemplified in the Impact section in which a lack of research into the specific "spokes" of the semantic cognition model was discussed. Furthermore, the need for increased research was reflected in the disparate areas of activation in the NeuroSynth Maps which showed only small

spots of activation, indicating the model may have required more data in order to fully flesh out the specific brain regions. Moreover, as noted by both Dockès et al. and Yarkoni et al., issues may arise due to the tools' lexical approach. For instance, the framework may not account for redundancies between terms, such as “semantic memory” and “lexical memory”.

Differences between the methodologies of NeuroSynth and NeuroQuery also affected the results of the study. While NeuroSynth uses text-mining of different studies' abstracts to generate maps (Yarkoni et al.), NeuroQuery is a model that uses prior data to predict brain activation. Thus, NeuroQuery may extend beyond the existing research environment to generate its maps, which would explain the differences between the activation maps as the NeuroQuery maps showed regions of activation past those from NeuroSynth. As strong levels of activation in the prefrontal cortex and left temporal lobe which supported the semantic cognition model were only shown in the NeuroQuery map, this emphasizes the need for more robust research that can confirm NeuroQuery's “predictions”.

Ultimately, the use of these tools must be supplemented by true, empirical research and other sources of evidence. As Dockès et al. note, Neuroquery cannot provide “conclusive evidence that a brain region is recruited by a mental process”. Therefore, more primary, controlled research must be conducted in order to fully confirm how semantic memory is encoded and stored.

Works Cited

- Binder, Jeffrey R., and Rutvik H. Desai. "The Neurobiology of Semantic Memory." *Trends in Cognitive Sciences*, vol. 15, no. 11, Nov. 2011, pp. 527–36, doi:10.1016/j.tics.2011.10.001.
- Borogovac, Ajna, and Iris Asllani. "Arterial Spin Labeling (ASL) FMRI: Advantages, Theoretical Constraints, and Experimental Challenges in Neurosciences." *International Journal of Biomedical Imaging*, vol. 2012, Feb. 2012, p. 818456, doi:10.1155/2012/818456.
- Bouyeure, Antoine, and Marion Noulhiane. "Memory: Normative Development of Memory Systems." *Handbook of Clinical Neurology*, vol. 173, 2020, pp. 201–13, doi:10.1016/B978-0-444-64150-2.00018-6.
- ChatGPT. Python code for neuroimaging statistical analysis using NiLearn. 2025. Python script.
- Chiou, Rocco, et al. "Controlled Semantic Cognition Relies upon Dynamic and Flexible Interactions between the Executive 'Semantic Control' and Hub-and-Spoke 'Semantic Representation' Systems." *Cortex*, vol. 103, June 2018, pp. 100–16, doi:10.1016/j.cortex.2018.02.018.
- Corballis, Michael C. "Left Brain, Right Brain: Facts and Fantasies." *PLoS Biology*, vol. 12, no. 1, Jan. 2014, p. e1001767, doi:10.1371/journal.pbio.1001767.
- Damasio, A. R. "Category-Related Recognition Defects as a Clue to the Neural Substrates of Knowledge." *Trends in Neurosciences*, vol. 13, no. 3, Mar. 1990, pp. 95–98, doi:10.1016/0166-2236(90)90184-c.
- Dockès, Jérôme, et al. "NeuroQuery, Comprehensive Meta-Analysis of Human Brain Mapping." *ELife*, vol. 9, Mar. 2020, doi:10.7554/eLife.53385.
- Eriksson, Johan, et al. "Neurocognitive Architecture of Working Memory." *Neuron*, vol. 88, no. 1, Oct. 2015, pp. 33–46, doi:10.1016/j.neuron.2015.09.020.
- Frisby, Saskia L., et al. "Decoding Semantic Representations in Mind and Brain." *Trends in Cognitive Sciences*, vol. 27, no. 3, Mar. 2023, pp. 258–81, doi:10.1016/j.tics.2022.12.006.
- Gainotti, Guido. "The Organization and Dissolution of Semantic-Conceptual Knowledge: Is the 'amodal Hub' the Only Plausible Model?" *Brain and Cognition*, vol. 75, no. 3, Apr. 2011, pp. 299–309, doi:10.1016/j.bandc.2010.12.001.
- Glover, Gary H. "Overview of Functional Magnetic Resonance Imaging." *Neurosurgery Clinics of North America*, vol. 22, no. 2, Apr. 2011, pp. 133–39, vii, doi:10.1016/j.nec.2010.11.001.
- Hickok, Gregory, and David Poeppel. "The Cortical Organization of Speech Processing." *Nature Reviews. Neuroscience*, vol. 8, no. 5, May 2007, pp. 393–402, doi:10.1038/nrn2113.
- Huber, Laurentius, et al. "High-Resolution CBV-FMRI Allows Mapping of Laminar Activity and Connectivity of Cortical Input and Output in Human M1." *Neuron*, vol. 96, no. 6, Dec. 2017, pp. 1253–1263.e7, doi:10.1016/j.neuron.2017.11.005.

- Lambon Ralph, Matthew A., et al. "The Neural and Computational Bases of Semantic Cognition." *Nature Reviews. Neuroscience*, vol. 18, no. 1, 2017, pp. 42–55, doi:10.1038/nrn.2016.150.
- Manns, Joseph R., et al. "Semantic Memory and the Human Hippocampus." *Neuron*, vol. 38, no. 1, Apr. 2003, pp. 127–33, doi:10.1016/s0896-6273(03)00146-6.
- Patterson, Karalyn, and Matthew A. Lambon Ralph. "The Hub-and-Spoke Hypothesis of Semantic Memory." *Neurobiology of Language*, Elsevier, 2016, pp. 765–75, doi:10.1016/B978-0-12-407794-2.00061-4.
- Patterson, Karalyn, et al. "Where Do You Know What You Know? The Representation of Semantic Knowledge in the Human Brain." *Nature Reviews. Neuroscience*, vol. 8, no. 12, Dec. 2007, pp. 976–87, doi:10.1038/nrn2277.
- Reilly, Jamie, and Jonathan E. Peelle. "Effects of Semantic Impairment on Language Processing in Semantic Dementia." *Seminars in Speech and Language*, vol. 29, no. 1, Feb. 2008, pp. 32–43, doi:10.1055/s-2008-1061623.
- Rugg, Michael D., and Kaia L. Vilberg. "Brain Networks Underlying Episodic Memory Retrieval." *Current Opinion in Neurobiology*, vol. 23, no. 2, Apr. 2013, pp. 255–60, doi:10.1016/j.conb.2012.11.005.
- Ryan, Lee, et al. "The Role of Medial Temporal Lobe in Retrieving Spatial and Nonspatial Relations from Episodic and Semantic Memory." *Hippocampus*, vol. 20, no. 1, Jan. 2010, pp. 11–18, doi:10.1002/hipo.20607.
- Squire, L. R., et al. "Amnesia: Declarative and Nondeclarative Memory." *Encyclopedia of Neuroscience*, Elsevier, 2009, pp. 289–94, doi:10.1016/B978-008045046-9.00548-9.
- Sridhar, Sruthi, et al. "Cognitive Neuroscience Perspective on Memory: Overview and Summary." *Frontiers in Human Neuroscience*, vol. 17, July 2023, p. 1217093, doi:10.3389/fnhum.2023.1217093.
- Thompson-Schill, Sharon L. "Neuroimaging Studies of Semantic Memory: Inferring 'How' from 'Where'." *Neuropsychologia*, vol. 41, no. 3, 2003, pp. 280–92, doi:10.1016/s0028-3932(02)00161-6.
- Tong, Jiaqing, et al. "A Distributed Network for Multimodal Experiential Representation of Concepts." *The Journal of Neuroscience*, vol. 42, no. 37, Sept. 2022, pp. 7121–30, doi:10.1523/JNEUROSCI.1243-21.2022.
- Vargha-Khadem, F., et al. "Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory." *Science*, vol. 277, no. 5324, July 1997, pp. 376–80, doi:10.1126/science.277.5324.376.
- Warrier, Catherine, et al. "Relating Structure to Function: Heschl's Gyrus and Acoustic Processing." *The Journal of Neuroscience*, vol. 29, no. 1, Jan. 2009, pp. 61–69, doi:10.1523/JNEUROSCI.3489-08.2009.
- Whitten, Lori A. *Functional Magnetic Resonance Imaging (Fmri): An Invaluable Tool in Translational Neuroscience*. RTI Press, 2012, doi:10.3768/rtipress.2012.op.0010.1212.

Wixted, John T., et al. “Coding of Episodic Memory in the Human Hippocampus.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 5, Jan. 2018, pp. 1093–98, doi:10.1073/pnas.1716443115.

Yarkoni, Tal, et al. “Large-Scale Automated Synthesis of Human Functional Neuroimaging Data.” *Nature Methods*, vol. 8, no. 8, June 2011, pp. 665–70, doi:10.1038/nmeth.1635.

Zhang, Yizhen, et al. “Connecting Concepts in the Brain by Mapping Cortical Representations of Semantic Relations.” *Nature Communications*, vol. 11, no. 1, Apr. 2020, p. 1877, doi:10.1038/s41467-020-15804-w.

Zlotnik, Gregorio, and Aaron Vansintjan. “Memory: An Extended Definition.” *Frontiers in Psychology*, vol. 10, Nov. 2019, p. 2523, doi:10.3389/fpsyg.2019.02523.

Appendix

Link to Python Notebook:

<https://colab.research.google.com/drive/1cFNIOyzjul59LWJ7tDWdmVSa-cGsGIIV?usp=sharing>

Analyzing the Russia-Ukraine Crisis's Impact on the Global Trade of Agricultural Products and Energy Resources By Syaon Kumar

Introduction

The exchange of goods beyond country borders has existed for thousands of years. We have imported and exported goods based upon the needs and wants of countries all over the world. One can think of this as exchanges that benefit all parties involved, an idea formulated by A. Smith in 1776. When it comes to the question of what to import and export, countries look to the resources and factors of production available to them. For a country like Ukraine, they may find an abundance of fertile land that could be used to grow crops, and thus may decide that the opportunity cost to produce agricultural products would be lower than that of others. To make up for the large production of agricultural products, Ukraine may engage in trade with another country with a comparative advantage in the production of machinery.

When countries take advantage of the underlying concepts of resource allocation and opportunity cost, they can optimize the flow of goods and services between each other to achieve higher levels of overall production. Using the idea of who can produce what at the highest efficiency, firms develop supply chains separating each step of a good's production process between various countries. However, as global trade routes strengthen and production grows, we risk becoming too dependent on each other for goods. For example, when the COVID-19 Pandemic arose in 2019, international trade severely decreased and industries took a hit as access to overseas goods grew limited.

Background of the Russia-Ukraine War

Russia remains a dominant military power with a large arsenal of nuclear warheads and the largest surface area in the world. Due to its landscape and resources, Russia sources a large amount of raw materials and energy through petroleum, gas, iron, and palladium (Darvas and Martins 28). In 1922, Ukraine joined the Soviet Union, then left the Union in 1991, then became one of the founding members of the Commonwealth of Independent States (Orhan 142). Due to the shared history as part of the Soviet Union, Russia finds Ukraine of special importance. Additionally, Ukraine's geopolitical position places the country at a high value to Russia.

The annexation of Crimea by Russia in 2014 enabled Russian forces to advance closer to Ukraine. Ukraine's location between Europe and Russia, its dominance over the Black Sea, and disputes over Crimea were all factors that incentivized Russia to invade Ukraine in 2022 (Orhan 141). Following the invasion, the United Nations passed two resolutions to demand Russia to reverse its attempted illegal annexation of Ukrainian land (Darvas and Martins 1). Refusing to adhere to the resolutions, Russia was imposed with economic sanctions disrupting global prices and markets. The aim of this paper is to consider various approaches to discussing the impact of the Russian-Ukrainian war on the international trade of energy resources and agricultural products and synthesize various data to answer this question.

Dependence on Russian and Ukrainian Goods

Russia has been a large exporter of energy resources like oil, natural gas, and minerals. Due to its large amount of land, Russia has vast amounts of oil and gas reserves as well as a natural supply of minerals. Because of this, many other countries around the world have engaged in trade with Russia to access these crucial goods. For transportation, automobile production, steel production, and battery manufacturing, Russian exports play a critical role in their manufacturing. Because of this, prior to the war, Russia supplied 19% of the world's natural gas and 11% of its oil (Orhan 144). For Europe, more than 50% of its thermal coal supply came from Russia, and European Union countries such as Cyprus, Lithuania, Greece, and Romania were completely dependent on Russian coal (Ziółkowski 24). Given the wide usage of Russian energy resources and the heavy dependence on Russian exports, the Russian-Ukraine crisis would harm the industries of various countries and force them to find suppliers elsewhere.

For Ukraine, there was much fertile land for agricultural products to be grown. Known as the chernozems of central Ukraine, this fertile soil covered over two-thirds of the country's area (Tass, Bhat, Riyaz, and Hakim 1148). This land has enabled Ukraine to specialize in the production of sunflower oil, wheat, barley, and corn (Darvas and Martins 8).

Together, Russian and Ukrainian wheat and sunflower products accounted for the great majority of the world's agri-food trade. These products were especially important for developing countries who may be reliant upon these imports. For example, over twenty African nations largely depend on the import of wheat from these two nations to feed their growing populations (Tass, Bhat, Riyaz, and Hakim 1148).

From a theoretical standpoint, an attack on Ukraine, combined with the trade sanctions later put on Russia, would create chaos as exports and imports would shift to meet the needs of the Ukrainian population. As one would expect, the effects of these trade shifts would not only impact the host countries, but would extend to overseas countries dependent upon Russian-Ukrainian goods. From there, these overseas countries would need to find other countries to substitute their imports such that production does not halt to a complete stop.

Consequences of the Russia-Ukraine Crisis on Energy and Agricultural Product Prices

The impact of the Russia-Ukraine crisis is evident through the change in prices of the commodities of these economics on the global scale. From January 2022 to February 24, 2022 (when the invasion occurred), there was a significant (>50%) increase in the prices of nickel and gas, and a moderate (10-25%) increase in the prices of oil, iron ore, palladium, gold, zinc, and copper. In terms of agricultural products, the prices of wheat and corn both rose significantly as well (Orhan 144).

In Orhan's study, this data was used to conclude that the disruptions of the Russia-Ukraine crisis caused profound economic and financial shocks, specifically in commodity markets where prices for energy resources have soared. For Europe, prices for gas would be ten times what they were a year ago, with the cost of oil doubling within the same period (Orhan 144). Given the high dependence of the European market for energy resources on Russian gas

and oil, this conflict disrupted Russian and Ukrainian production, causing the previously mentioned commodity prices to skyrocket in those respective markets (Orhan 142).

It is important to note that this approach may lack in taking into account how much of the increase in prices since January were actually a cause of the Russia-Ukraine war. Orhan's study's approach relies on comparisons between prices from January of 2022, which was before the invasion of Ukraine, to February, 24 and mid-March. In his approach, Orhan's analysis does not contrast price increases from January 2022 to February 24, 2022 with price increases from January 2022 to mid-March 2022, possibly overlooking other factors that could be in play here. In fact, for agricultural products, the data from Orhan's paper directly shows that for the price of wheat, the percent change from January 2022 to February 24, 2022 was roughly 88%, while the percent change from January 2022 to mid-March 2022 was only approximately a percent higher than that (89%). For the price of corn, the percent change from January 2022 to February 24, 2022 was ~43%, whereas the percent change from January 2022 to mid-March 2022 was ~49% (Orhan 144). This means that for agricultural products like wheat and corn, the price of these commodities had risen significantly more before the war had started on February 24, 2022, than it had risen after the war to March of 2022. The sharp increase of prices before the war conflict even started indicates there are other factors at play than the disruptions caused by the war.

Furthermore, we can also observe comparisons in prices across the various time periods for energy resources to indicate that war was not the main driver of global prices. For the price of coal, the percent change from January 2022 to February 24, 2022 was roughly 80%, while the percent change from January 2022 to mid-March 2022 was roughly 90% (Orhan 144). For the price of Gas Europe, Oil, Palladium, Iron Ore, and Aluminum, the percent change from January 2022 to February 24, 2022 was larger than the percent change from January 2022 to mid-March 2022 (Orhan 144). From this data, we can conclude that global prices for energy resources changed more profoundly before the war conflict started, as the change from February 24, 2022 to mid-March 2022 was far less than the change from January 2022 to February 24, 2022.

Due to these observations, Orhan's approach to this topic may capture the Russia-Ukraine crisis as a factor for the increase of global prices of both energy resources and agricultural products; however, it may fail to take into consideration that prices had already been increasing before the war and thus, it may be difficult to attribute overall soaring prices of energy resources to the disruptions of the Russia-Ukraine crisis.

The impact of this crisis can also be observed through analysis of data on the percent changes of forecast prices to conclude that the war and sanctions (placed upon Russia) were not the most important drivers of global energy and agricultural prices, with other elements at hand (Darvas and Martins 10).

To capture the general trend, Darvas and Martins use the International Monetary Fund forecast revisions from October 2021 to October 2022 to state that during the time period a little before and after the war started, energy forecast prices increased significantly and non-energy commodities increased moderately. Using information on the trading prices of Brent Crude Oil and Dutch TTF gas prices, they found that prior to the war, both prices increased 54% and 680%

respectively, then shortly after the invasion, both prices fluctuated before returning to their value the year before.

Darvas and Martins' analysis synthesizes various data to arrive at the conclusion that in the year before Russia invaded Ukraine, commodity prices had increased by 52%, faster than in any year since when the data became available in 1992. Using the idea that energy prices increased faster in 2021 than in 2022, they suggested that war was not the main driver of energy prices.

As for agricultural products, Darvas and Martins analyzed food price forecasts by the International Monetary Fund for wheat, barley, and sunflower products, noting an average increase by 11% for these forecasts. While it is clear that commodity prices rose during this time period, the researchers found it difficult to isolate what event(s) were the main driving force behind price increases before the invasion of Ukraine and after the invasion of Ukraine.

It is important to note that before the war, global inflation pressures had been building up with the prevalence of the COVID-19 pandemic (Darvas and Martins 1). This global outbreak resulted in shortages as transportation costs surged and countries were forced to limit contact with one another. When limitations eased up in 2021, demand for goods soared resulting in a global scarcity of goods, driving inflation (Darvas and Martins 1). During this period, Russia had already begun reducing natural gas exports, and combined with shutdowns in nuclear power plants in France (increasing the need for gas in power generation) along with severe droughts which increased demand for liquefied natural gas, the price of energy resources like gas shot up even higher (Darvas and Martins 2).

This approach does account for the fact that prices had already been increasing before the war, yet pinpointing exactly what changes in prices were solely a result of this crisis is difficult, especially with other previously mentioned factors like the COVID-19 pandemic and the shutdown of nuclear power plants in France also prevalent during this time period. Still, it is clear that shortly after the war, prices of energy resources and agricultural products increased, likely due to disruptions in exports and imports that led to a decrease in overall supply of these resources, yet in the long-run, prices settled back to their pre-war values as countries sought new countries to export to and import from.

Shifts in patterns of trade as a result of the Russia-Ukraine War

Another way to analyze the impact of the Russia-Ukraine crisis is to examine how disruptions in supply chains altered global trade routes. These disruptions would stem from three sources: restricted transit through Russia, closed airspace between the EU countries and Russia, and cancelled sea freight routes through the Black Sea (Orhan 145).

From a safety perspective, continuing trading through Russia could pose risks as condemnation of Russia's actions by the EU could turn Russia hostile. Land-based routes between Europe and Asia would thus become limited, resulting in decreased connection between European countries and Asian countries from which trade had grown to be strong. Goods for which European countries, like Ukraine, specialized in would face barriers in getting exported to

countries which need them, contributing to global scarcity and shortages in goods like staple foods, directly contributing to a rise in prices.

The decision of the European Union countries to close their airspace to Russia aircraft and cargo as a safety measure would result in hampered ties between Russia and 36 other countries (Orhan 145). Closing off ties would mean that Russia would no longer be able to export gas and oil to EU countries, contributing to scarcity. This is yet another factor contributing to rising prices as there is less supply of crucial energy resources on the market, meaning whatever is able to be supplied faces an increase in monetary value as countries attempt to get their hands on what remains.

Following the invasion of Ukraine, they were forced to shut down commercial shipping to focus on the war effort (Orhan 145). This would result in a halt of sea freight routes through the Black Sea due to the close proximity to Russia and Ukraine that would pose security risks to transportation, and there would be less Ukrainian focus on these routes (Orhan 145). Thus, not only would land routes between Europe and Asia take a hit, but former sea freight routes would be forced to close over security concerns and sanctions imposed on Russia. This war would end up negatively impacting global cargo capacity due to decreased trade routes and further limitations, and would increase freight prices as freight forwarders would need to take longer routes to avoid contact with the Russia-Ukraine crisis, resulting in more fuel consumption and a lower frequency of trade (Orhan 145).

If we take a look at the Ukrainian scene, it is evident that Ukraine's production costs would face an increase. As one would expect, an invasion by Russia and further military intervention would damage Ukrainian infrastructure as well as fields. This damage would have the negative effect of decreasing Ukraine's ability to produce and export wheats, barley, corn, and other products as there would be fewer factors of production to take advantage of. Conscription in Ukraine paralleled with fleeing refugees would mean a loss of labor supplied for harvesting crops and preparing them to be exported. All these factors would pose hardships for Ukraine to maintain its level of exports, hampering the ability of countries to trade with Ukraine and forcing them to find other suppliers.

Combined, all three sources would result in a lower frequency of trade and a greater scarcity of energy resources and agricultural products. EU countries would not be able to import oil and gas from Russia, and countries around the world that rely on Ukraine's agricultural products would need to find alternatives as Ukraine would be focused on the war at hand. From this logic, one would expect prices to rise to some extent due to unchanged demand but decreased supply, which would match the increased prices established by both papers in the previous section.

In order to examine the effects of these supply chain disruptions, we will first focus on how this conflict impacted Russia. Following the invasion, the United States and European Union imposed many economic sanctions on Russia to discontinue relations (Darvas and Martins 1). These sanctions prohibited exports of strategic goods to Russia (high-tech goods and electronic components) as well as imports of oil and gas from Russia (Darvas and Martins 1).

As a result of these sanctions, Russia was forced to find other countries to export their energy resources to, resulting in a total trade surplus with the United States falling from a peak of 2.5 billion to below 1 billion and a total trade surplus with the United Kingdom falling from a peak of 2.2 billion to below 0.5 billion (Darvas and Martins 15). To compensate, from January 2022 to May of 2022, Russia increased its exports in all of their product categories to India, China, and Turkey, decreasing exports to the United States and United Kingdom (Darvas and Martins 15). This is due to the fact that these three countries did not join the United States nor the United Kingdom in imposing sanctions on Russia as a response to the invasion and instead chose to remain neutral, resulting in stronger trade relationships with Russia.

Overall, while sanctions did prevent Russia from exporting to countries aligned with the United States and the United Kingdom, Russia's total trade surplus from before and after the start of the war rose over ten billion, mainly consisting of mineral fuels (Darvas and Martins 13). Russia shifted its trade patterns from relations with Western countries to stronger relations with Eastern countries instead, demonstrating how Russia was able to continue profiting off its abundance of oils and gases even through the sanctions.

As for other countries, they were forced to find other suppliers of the goods they desired and needed for their economy. In the European scene, Cyprus, Lithuania, Greece, and Romania all previously imported energy resources from Russia, but now reoriented their imports to main suppliers from Columbia, Australia, and South Africa (Ziółkowski 24). Other European nations like Poland and Germany found that they needed to become less reliant on others and boost independence, leading to Poland increasing its domestic production and Germany reopening closed mines to make up for the loss of imports of energy resources (Ziółkowski 24).

As Ukraine was forced to decrease exports of agricultural products, a shortage of staple foods in the global market developed. For countries in Africa and the Middle East previously importing cereals and food from Ukraine, they needed to find alternatives to cater to the domestic demand for these products (Sharma, Ankit, Mittal, Ravi, and Jarial 64). Patterns of trade shifted as seen by a lower frequency of trade between developing nations and Ukraine, which was replaced by greater trade between developing nations and the United States and the United Kingdom as well as greater focus on domestic production of wheat and cereals (Ziółkowski 25).

Although these nations were able to adapt to the consequences of the Russian invasion of Ukraine, not only were food prices much higher due to shortages, but an increased focus on domestic production meant more labor would be required for farming jobs, decreasing the amount of labor available for other sectors like manufacturing and technologies. This crisis temporarily halted advancement in developing countries as they used the resources they had, including labor and capital, on the agricultural industry to first ensure the survival of their populations.

Conclusion

In analyzing the overall effects of the Russia-Ukraine crisis on the international trade of energy resources and agricultural products, it is important to consider both changing prices within markets as well as shifting trading patterns. Although prices of these commodities had already been rising on the global scale before this conflict began, it is evident that this conflict further contributed to the surge in prices in the short-term due to the observable decrease in exports from Ukraine and shifts in exports from Russia. As for the long-term, countries developed new shipping routes such that they could continue to profit off of their natural resources (Russia), while countries previously dependent on Ukraine relied on imports from the United States and the United Kingdom to sustain their growing population.

The idea of international trade dates back to the development of the Silk Roads, where varying empires sought to get their hands on prized possessions from different areas of the Eastern Hemisphere, including valuable silk and porcelain from China. Across history, we have developed numerous methods of transportation, ranging from cargo ships to railways. These developments have enabled us to connect with countries on opposite sides of the world, allowing us to build a web of connections connecting all continents and almost all countries to one another.

This is especially important when it comes to creating supply chains, as the production of many goods involves dozens of countries working together in harmony—sourcing raw materials, converting them into parts, assembling those components elsewhere, and then finally selling the finished product in yet another location. Without the trade of these resources between countries, we would not have been able to construct complex supply chains that exist to this day and help us to achieve efficiency in production.

As this crisis continues to this day, although there are mentions of peace talks and resolutions in the near future, it is unclear when this war may truly come to an end. As the war has persisted for almost three years now, countries previously reliant upon Ukrainian exports have now found alternatives for the supply of agricultural products. In Africa, they increased focus on national production, coupled with imports from nearby outside regions to replace the agricultural products they previously imported from Ukraine. Due to years of increased focus on this, Africa may continue to increase national production of agricultural products such that they do not need to rely on other countries for resources that their population requires for survival.

As for the trade routes that either developed or strengthened over this time period, they would continue to operate as countries have adapted to new imports that they used to substitute the old imports they received. Even if the war effort ended in the near future, it would take Ukraine a few years before it is able to rebuild its infrastructure, restart its shipping routes, gather a labor force, and stabilize the economy to the point where it can participate on the global market again, a difficult recovery given its state as of now.

Works Cited

- Darvas, Zsolt, and Catarina Martins. "The Impact of the Ukraine Crisis on International Trade." Bruegel, 27 Aug. 2025, www.bruegel.org/working-paper/impact-ukraine-crisis-international-trade.
- Orhan, Ebru. "THE EFFECTS OF THE RUSSIA - UKRAINE WAR ON GLOBAL TRADE." *Journal of International Trade, Logistics and Law*, vol. 8, no. Number 1, 1 June 2022, pp. 141–146.
- Sharma, Kushagra, et al. "THE RUSSIA-UKRAINE CONFLICT: IMPLICATIONS FOR GLOBAL FOOD SECURITY, TRADE, AND ECONOMIC STABILITY." *Food and Agri Economics Review*, vol. 4, no. 2, 28 May 2024, pp. 39–45, <https://doi.org/http://doi.org/10.26480/faer.02.2024.39.45>.
- Tass, Mudasir Ahmad, et al. "Russia-ukraine war and its impact on global oil and food prices." *Journal of Advanced Zoology*, vol. 45, no. 1, 17 Jan. 2024, <https://doi.org/10.53555/jaz.v45is1.3556>.
- Ziółkowski, Krzysztof. "The impact of the Russia-Ukraine conflict on World Trade." *The Business and Management Review*, vol. 14, no. 01, 10 Apr. 2023, <https://doi.org/10.24052/bmr/v14nu01/art-03>.

Exploring the Relationship Between United States Electric Vehicle Demand and Taiwanese Manufacture of Integrated Circuits By Joshua Liu

Abstract

In this paper, we study whether monthly U.S. plug-in electric-vehicle demand is associated with Taiwan's integrated-circuit wafer fabrication output. Using monthly data from Dec. 2010 to June 2025, we retrieve fabrication data from the Taiwan Ministry of Economic Affairs (MOEA); specifically, foundry wafers aggregated into a 12-inch equivalent series. EV demand is quantified with Argonne National Laboratory's monthly PEV totals.

An autoregressive distributed-lag model (IC log on its own lags and PEV log lags L0–L12) fits well ($R^2 = 0.959652$; $SE_\gamma = 0.049348$; $F = 130.338$; $n = 163$; $df_{\text{resid}} = 137$). We find a small, immediate positive association ($\beta_0 = 0.0418$, $p = 0.0427$), a negative response at two months ($\beta_2 = -0.0671$, $p = 0.0056$), and a near-zero cumulative elasticity (sum of EV lag coefficients L0-L12) over 13 months ($\Sigma\beta = 0.004637$). A partial F-test ($F \approx 1.81$ with 13, 137 df) shows the EV block adds only modest explanatory power beyond IC persistence. Taken together, the pattern suggests that PEV demand shifts the timing of IC output (a contemporaneous bump and short-lag give-back) rather than producing a large, sustained increase over a one-year horizon. We release the underlying workbook and README via a DOI for replication, noted in the Methods and Works Cited sections.

Introduction

Semiconductors—especially advanced ICs fabricated on 300 mm, 200 mm, and 150 mm wafer lines—are a core input to vehicle electrification. Policymakers track plug-in adoption closely, while manufacturers manage capacity and lead times that unfold over months, not days. Taiwan's foundries sit at the center of this landscape, yet there is limited, high-frequency evidence on how short-run changes in U.S. plug-in demand relate to Taiwanese wafer fabrication as it is actually measured in official statistics.

We ask a simple question: When U.S. PEV sales change, do we observe a contemporaneous or lagged change in Taiwan's IC wafer output, and of what magnitude? To answer it, we construct a monthly panel that (i) focuses narrowly on fabrication throughput—MOEA's (2611) Manufacture of Integrated Circuits foundry wafer items 2611110, 2611120, and 2611130—combined into a 12-inch-equivalent series; and (ii) pairs it with Argonne's monthly PEV totals. This approach anchors the outcome in finished wafer counts (not design services, masks, or substrate inputs), pairs it with a widely tracked monthly demand series, and states the timing assumptions explicitly.

We preview three findings. First, there is a small, statistically significant contemporaneous elasticity of IC output with respect to PEV demand. Second, there is a negative adjustment two months later. Third, the net 13-month effect is near zero, and the PEV block adds only modest explanatory power relative to IC's own momentum. These results are consistent

with capacity and scheduling frictions: demand shocks can pull production forward, but much of the move is subsequently unwound.

Literature Review

The International Energy Agency's Global EV Outlook 2025 reports that electric-car sales exceeded 17 million in 2024 and topped 20% of global market share, making EV demand a central lever for climate, energy security, and industrial competitiveness.

For month-to-month U.S. demand monitoring, Argonne National Laboratory maintains a monthly plug-in EV (PEV) series used widely by analysts and agencies; the Bureau of Transportation Statistics explicitly references Argonne's tables in its official materials, underscoring that this is the federal barometer for adoption.

U.S. demand-side policy shapes that signal directly: the IRS's Clean Vehicle Credit offers up to \$7,500 with income, price, assembly, and battery-content rules—meaning tax design influences who buys which EV and when.

On the supply side, fabrication capacity remains concentrated. TrendForce's 2Q25 round-up puts TSMC near 70% global foundry market share, amplifying how U.S. demand can propagate to Taiwanese fabs and why resilience concerns persist. The Semiconductor Industry Association's 2025 State of the U.S. Semiconductor Industry adds that announced private-sector investments now exceed half a trillion dollars, setting the baseline for CHIPS-era capacity and workforce efforts.

Policy relevance also hinges on operational frictions. TechInsights reports automotive semiconductor lead times up to ~12.9 weeks—still multi-month—so incentives may shift the timing of output rather than instantly expanding totals, consistent with the short-run bump and subsequent normalization we observe.

Finally, the federal supply-side response is formalized through CHIPS for America; NIST's CHIPS portal outlines the incentive architecture and ongoing program implementation—critical complements to demand-side EV policy if the goal is resilient, domestic (or allied) semiconductor supply supporting transportation electrification.

Together, these sources show why linking EV demand to IC fabrication matters for policy: EV uptake is accelerating (IEA); agencies rely on timely demand indicators (Argonne, BTS); incentives steer near-term purchasing (IRS); foundry capacity is concentrated (TrendForce) even as the U.S. scales investment (SIA); lead times remain meaningful (TechInsights); and CHIPS provides the structural supply response (NIST).

Methods

We assemble a monthly dataset combining U.S. plug-in electric-vehicle (PEV = BEV + PHEV) sales from Argonne National Laboratory's Light-Duty Electric Drive Vehicles Monthly Sales Updates with Taiwan's Ministry of Economic Affairs (MOEA) product-level statistics under (2611) Manufacture of Integrated Circuits. Argonne provides a consistent BEV/PHEV

time series; MOEA publishes product counts and units for specific IC items in both Chinese and English.

We restrict our analysis to plug-in electric vehicles (PEVs)—i.e., battery-electric (BEV) and plug-in hybrid (PHEV)—and treat non-plug-in hybrids (HEVs) separately because PEVs are grid-charged and therefore embody larger batteries and high-voltage power-electronics subsystems (traction inverter, onboard charger, DC-DC converter) that drive the semiconductor demand we aim to proxy with foundry wafer output (Alternative Fuels Data Center). This focus also improves measurement consistency: Argonne publishes a clean monthly PEV series used widely in U.S. tracking, enabling us to align demand timing with our monthly IC production measure. Finally, as the International Energy Agency describes, most policy targets and market outlooks emphasize plug-ins (BEV+PHEV) as the core electrification category, so concentrating on PEVs maximizes relevance to the policy and industry context we discuss.

We conducted the analysis in three steps. First, we pulled monthly PEV totals from Argonne National Laboratory's Light-Duty Electric Drive Vehicles Monthly Sales Updates and monthly IC product counts from Taiwan's Ministry of Economic Affairs (MOEA) for (2611) Manufacture of Integrated Circuits. From MOEA we kept the foundry wafer items 2611110 (12"+), 2611120 (8"), and 2611130 (6" and below) and combined them into the 12-inch-equivalent series described earlier. Next, we took natural logs—IC_log as $\ln(12\text{-inch-equivalent})$ and EV_log as $\ln(\text{PEV} + 1)$ to handle early zero months—and built monthly lags (EV_log from L0 to L12, IC_log from L1 to L12). Finally, we ran a regression where current IC_log is explained by its own past values and by current and past EV_log, and we reported the usual outputs (coefficients, standard errors, p-values, R^2) plus the sum of the EV effects across all lags. As a check, we also ran a version without the EV terms and compared the errors to see what the EV block adds. The raw MOEA series are kept as-is; the constructed data and results are saved in the workbook and archived with a README at our dataset DOI. Data and replication files (RESEARCH_DATA.xlsx and README) are openly available at Zenodo, version 1.0, DOI: 10.5281/zenodo.17167345.

Results

We report four key findings. We estimate a monthly distributed-lag model of Taiwan's IC wafer output on its own lags and on U.S. PEV demand. The model fits the data well ($R^2 = 0.959652$; $SE_\gamma = 0.049348$; $F = 130.338$; $n = 163$; $df_{\text{resid}} = 137$). IC production is highly persistent: past IC levels explain much of current output, which is typical for fabrication series.

Turning to U.S. PEV demand, the contemporaneous coefficient (L0) is positive and statistically significant: $\beta_0 = 0.042$ ($SE = 0.020$; $t = 2.045$; $p = 0.043$). Interpreted as an elasticity, a 1% increase in PEV sales this month is associated with about a 0.042% higher IC wafer output in the same month, holding IC's own lags constant. Most other individual PEV lags are small and not statistically distinguishable from zero; a notable exception is at L2, $\beta_2 = -0.067$ ($SE = 0.024$; $t = -2.813$; $p = 0.006$), consistent with a short-run give-back about two months after the initial uptick. One mid-lag is borderline (L5 = 0.042; $p = 0.067$).

Summing the EV-lag coefficients over the 13-month window yields a cumulative EV elasticity of 0.005 (L0-L12, 13 months), indicating that the long-run (net) effect is near zero. In other words, we see a small, immediate bump followed by a partial reversal, with little net change over a year.

Evaluated jointly, the EV lags provide only modest incremental explanatory power beyond IC's own persistence: the unrestricted model's SSE = 0.334 versus 0.391 for the IC-only restricted model, yielding a partial $F \approx 1.808$ with (13, 137) df.

Overall, the pattern is clear: strong IC persistence, a small contemporaneous positive association between PEV demand and IC wafer output, a short-lag negative offset, and very little long-run effect over 13 months.

Discussion

The short-run bump is plausible given the steady, widely tracked monthly PEV series maintained by Argonne National Laboratory, which offers a clean, high-frequency demand barometer for the U.S. market and is commonly used in policy and industry tracking. At the same time, the muted long-run elasticity is consistent with how advanced fabrication plants operate: TSMC's 2024 Annual Report emphasizes capacity in 12-inch-equivalent wafers and shows that additions occur on multi-year horizons, so monthly demand shocks often change timing more than total output. The lead-time frictions seen during the 2021–22 crunch—where average chip delivery times stretched to roughly ~20 weeks—also fit the pattern we estimate: orders can be pulled forward (raising contemporaneous output) and then normalize as scheduling catches up, leaving little net change after several months (Data Center Knowledge). Our dependent variable is intentionally narrow—Taiwan MOEA's (2611) Manufacture of Integrated Circuits foundry wafer items—so we are measuring fabrication throughput rather than design services, masks, or substrate materials; MOEA's item definitions make this distinction explicit and justify our scope. Finally, EV demand is global: the IEA's Global EV Outlook 2024 documents rapid worldwide adoption and rising shares, suggesting that future work could pair global (or regional) EV indicators with technology-specific fab metrics to test whether broader demand proxies yield stronger long-run signals.

Conclusion

We find that monthly U.S. plug-in EV demand is detectable in Taiwan's IC wafer fabrication data immediately ($\beta_0 \approx 0.042$) but that this bump is partly reversed two months later ($\beta_2 \approx -0.067$), leaving a near-zero cumulative elasticity across a 13-month window. The overall fit of the ADL model is strong because IC output is highly persistent; however, the incremental contribution of the PEV lags is modest at conventional thresholds. Overall, short-run EV demand shocks primarily re-time wafer output rather than expanding it over a one-year horizon.

Policy and management implications follow. Demand-side incentives can move near-term orders and utilization, but sustained increases in fabrication likely require complementary supply-side responses (capacity, tooling, workforce) or shifts in product/technology mix.

Measurement matters: by focusing on foundry wafer counts, we obtain a clean fabrication signal, but we do not capture packaging/test or non-IC segments that may respond differently.

Limitations include the U.S.-only demand proxy, reduced-form identification, and our fabrication-only outcome. Future work could (i) incorporate global or regional EV indicators, (ii) disaggregate by technology/node or by power devices, (iii) test robustness with alternative error structures and window lengths, and (iv) explore structural designs that separate demand shocks from supply constraints. We provide a DOI with the exact workbook and README to facilitate replication and extension.

Works Cited

- Alternative Fuels Data Center. “Electric Vehicles.” Alternative Fuels Data Center, U.S. Department of Energy, <https://afdc.energy.gov/vehicles/electric>. Accessed 15 Sept. 2025.
- Argonne National Laboratory. “Light-Duty Electric Drive Vehicles Monthly Sales Updates.” 2025, https://www.anl.gov/sites/www/files/2025-09/Total%20Sales%20for%20Website_August%202025.pdf. Accessed 13 Sept. 2025.
- Bloomberg. “Chip Delivery Time Surpasses 20 Weeks in No Sign Shortage Easing.” Data Center Knowledge, 10 Aug. 2021, <https://www.datacenterknowledge.com/data-center-chips/chip-delivery-time-surpasses-20-weeks-in-no-sign-shortage-easing>. Accessed 18 Sept. 2025.
- Bureau of Transportation Statistics. “Hybrid-Electric, Plug-in Hybrid-Electric and Electric Vehicle Sales.” 2025, <https://www.bts.gov/content/gasoline-hybrid-and-electric-vehicle-sales>. Accessed 18 Sept. 2025.
- Internal Revenue Service. “Credits for New Clean Vehicles Purchased in 2023 or After.” 3 Aug. 2025, <https://www.irs.gov/credits-deductions/credits-for-new-clean-vehicles-purchased-in-2023-or-after>. Accessed 19 Sept. 2025.
- International Energy Agency. Global EV Outlook 2024: Moving towards Increased Affordability. Apr. 2024, <https://www.iea.org/reports/global-ev-outlook-2024>. Accessed 3 Sept. 2025.
- International Energy Agency. Global EV Outlook 2025. 14 May 2025, <https://www.iea.org/reports/global-ev-outlook-2025>. Accessed 19 Sept. 2025.
- Liu, Joshua. EV Demand & Taiwan IC Foundry Output: Research Data and README. Zenodo, 19 Sept. 2025, <https://doi.org/10.5281/zenodo.17167345>.
- National Institute of Standards and Technology. “CHIPS for America.” 2025, <https://www.nist.gov/chips>. Accessed 19 Sept. 2025.
- Semiconductor Industry Association. 2025 State of the U.S. Semiconductor Industry. 7 July 2025, <https://www.semiconductors.org/wp-content/uploads/2025/07/SIA-State-of-the-Industry-Report-2025.pdf>. Accessed 19 Sept. 2025.
- Taiwan, Ministry of Economic Affairs. “C-2 Production, Sales, and Inventory Values of Major Products in the Information and Electronics Industry—By Product Type.” MOEA Department of Statistics, https://www.moea.gov.tw/mns/DOS/content/wHandMenuFile.ashx?file_id=36246. Accessed 16 Sept. 2025.
- Taiwan, Ministry of Economic Affairs. “Industrial Product Classification.” MOEA Department of Statistics, https://www.moea.gov.tw/mns/DOS/content/wHandMenuFile.ashx?file_id=27104. Accessed 13 Sept. 2025.

- Taiwan, Ministry of Economic Affairs. “Industrial Production Statistics (2611: Manufacture of Integrated Circuits).” 2025,
<https://service.moea.gov.tw/EE520/investigate/InvestigateDA.aspx?lang=E>. Accessed 12 Sept. 2025.
- Taiwan, Ministry of Economic Affairs. “IV. Appendix—Comparison Table of Industrial Product Classification Before and After Revision.” n.d.,
https://www.moea.gov.tw/Mns/DOS/content/wHandMenuFile.ashx?file_id=27105. Accessed 18 Sept. 2025.
- Taiwan Semiconductor Manufacturing Company. Annual Report 2024. 28 Feb. 2025,
https://investor.tsmc.com/sites/ir/annual-report/2024/2024%20Annual%20Report_E.pdf. Accessed 19 Sept. 2025.
- TechInsights. “Market Model: Automotive Semiconductor Lead Times for Commodity Components 2025 Q1.” 23 May 2025,
<https://www.techinsights.com/blog/market-model-automotive-semiconductor-lead-times-commodity-components-2025-q1>. Accessed 18 Sept. 2025.
- TrendForce. “2Q25 Foundry Revenue Surges 14.6% to Record High, TSMC’s Market Share Hits 70%.” TrendForce Press Center, 1 Sept. 2025,
<https://www.trendforce.com/presscenter/news/20250901-12691.html>. Accessed 19 Sept. 2025.

Sociodemographic Predictors of Delayed or Unmet Medical Care Among California Adolescents

By Lauren Chen

Abstract

Delayed or unmet medical care among adolescents is a critical public health concern. Delayed or unmet medical care occurs when adolescents either postpone or entirely go without needed health services, regardless of the cause. Although we know the prevalence of missed, skipped, or delayed preventive checkups due to the COVID-19 pandemic, we do not know the prevalence of delayed or unmet medical care in general following the COVID-19 pandemic. We also do not know the socio-demographic risk factors associated with delayed or unmet medical care among adolescents. We hypothesize that adolescents with greater socio-demographic vulnerability (lower income, immigrant background, larger family size, or unstable insurance coverage) are more likely to experience delayed or unmet medical care, and that the prevalence of delayed or unmet care differs significantly across age, gender, and racial/ethnic groups. We utilized the 2023 California Health Interview Survey, which sampled 1,045 adolescents aged 12 to 17 years old. We fit a weighted logistic regression model with delayed or not receiving medical care as the outcome of interest. Covariates included age, self-reported gender, race/ethnicity, citizenship status, rural/urban residence, family size, difficulty getting by on family income, and health insurance status in the past 12 months. Our central findings include the adjusted odds ratios and confidence intervals. In conclusion, delayed or unmet medical care among California adolescents remains a significant and inequitable challenge, with elevated risks among females, racial or ethnic minority youth, and those with unstable health insurance. Future research could build on this foundation by conducting longitudinal studies to examine the downstream health, developmental, and educational consequences of unmet medical care. By addressing structural and policy barriers, California has the opportunity to improve adolescent health outcomes and set a model for other states seeking to ensure equitable access to care.

1 Introduction

Delayed or unmet medical care among adolescents is a critical public health concern¹⁻³. A nationally representative study found that 19.2% of adolescents reported unmet health care need, which independently predicted worse adult health outcomes, including fair or poor general health, functional impairment, depressive symptoms, and suicidal ideation⁴. This prevalence highlights a substantial burden given the unique health needs of adolescents and the potential for both short and long-term consequences⁵. Delayed or unmet medical care occurs when adolescents either postpone or entirely go without needed health services, regardless of the cause⁶.

The risks associated with delayed or unmet medical care are multifaceted. Infectious diseases represent a particularly acute threat⁷. For example, failure to receive timely influenza vaccinations has been associated with increased risks of contracting influenza⁸. The influenza vaccine has been shown to reduce pediatric mortality by about 65% compared to unvaccinated

children⁹. Other infections, such as mononucleosis and meningitis, may also go undetected or untreated, resulting in prolonged illness or severe complications^{10,11}. Beyond acute infections, delayed care can hinder the management of chronic conditions such as asthma, diabetes, and obesity, exacerbating morbidity and increasing the risk of hospitalization¹². Missed diagnoses also contribute to poor outcomes, as adolescents may present later with more advanced disease stages, reducing treatment effectiveness¹³.

Psychological and developmental outcomes are also at risk. Delayed access to mental health services can worsen depression, anxiety, and behavioral disorders¹⁴. Additionally, inadequate pediatric care has been linked to delayed school enrollment, missed educational opportunities, and long-term developmental consequences¹⁵.

Multiple factors contribute to why adolescents delay or forgo medical care. Financial barriers, including high out-of-pocket costs and inadequate insurance coverage, remain major factors¹⁶. Insurance instability, such as being uninsured for part of the year, further increases vulnerability¹⁷. Structural barriers, including transportation difficulties, lack of access to routine primary care, and shortages of healthcare workers (e.g., physicians, nurse practitioners, physician assistants, and nurses), also limit timely care^{18,19}. Together, these factors highlight the need to better understand which groups of adolescents are at higher risk of unmet medical care and the consequences that may follow.

Wisk et al. (2012) found that nearly 10% of US families did not receive needed healthcare between 2001 and 2008²⁰. During the early period of the COVID-19 pandemic, Anderson et al. (2021) found that approximately 4 in 10 adults reported forgoing medical care²¹. Lebrun-Harris et al. (2021) utilized data from the US Census Bureau's Household Pulse Survey, collected in April and May 2021, and found that 26.4% of adolescents nationally missed or delayed preventive visits due to the COVID-19 pandemic²². Lebrun-Harris et al. (2021) also found substantial variation in the prevalence of missed or delayed preventive visits across states: from 17.9% in Wyoming to 37.0% in Vermont. Notably, the prevalence was 29.9% in California, 28.6% in Texas, and 23.4% in New York.

Although we know the prevalence of missed, skipped, or delayed preventive checkups due to the COVID-19 pandemic, we do not know the prevalence of delayed or unmet medical care in general following the COVID-19 pandemic. We also do not know the socio-demographic risk factors associated with delayed or unmet medical care among adolescents. We hypothesize that adolescents with greater socio-demographic vulnerability (lower income, immigrant background, larger family size, or unstable insurance coverage) are more likely to experience delayed or unmet medical care, and that the prevalence of delayed or unmet care differs significantly across age, gender, and racial/ethnic groups.

2. Methods

2.1 Data.

The 2023 California Health Interview Survey (CHIS) employed an address-based sample design to represent the state's non-institutionalized household population, ensuring estimates at both local and statewide levels across major racial, ethnic, and smaller subgroups²³. Sampling was stratified across California's 58 counties, with additional sub-strata in Los Angeles and San Diego to capture geographic diversity, and interviews were conducted in six languages (English, Spanish, Mandarin, Cantonese, Vietnamese, Korean, and Tagalog) to reach non-English speakers. Within selected households, one adult was randomly chosen, and when applicable, one adolescent and one child were also included, with adolescents interviewed directly and children represented by their parent or guardian. To generate population-level estimates, survey weights adjusted for selection probabilities and design features. CHIS data are widely used by federal, state, and local agencies, healthcare providers, and researchers to assess health needs, guide policy development, allocate resources, and advance public health research.

The outcome of interest was delayed or unmet medical care, measured by the survey question: "During the past 12 months, was there any time when you delayed or did not receive medical care that you thought you needed?" (Yes/No). Covariates included age (years), self-reported gender (male, female), race/ethnicity (Latino, Non-Latino Asian, Non-Latino Other [1 race], Non-Latino Other [2+ races], Non-Latino White), citizenship status (naturalized/non-citizen, US-born citizen), rural/urban zip code, family size (2-3, 4, 5+), how often it is hard to get by on family's income (never, not very often, somewhat often, very often), and insurance status in the past 12 months (insured all year, uninsured all year, uninsured part year).

CHIS 2023 sampling weights were applied to produce population-representative estimates that account for the survey's complex design. Weights adjust for the probability of selection at the household and individual levels, nonresponse, and post stratification to align with known population totals from sources such as the American Community Survey²⁴. Separate weights were calculated for adults, adolescents, and children, ensuring that estimates reflect California's non-institutionalized population overall, as well as those for key demographic and geographic subgroups. Weighted analyses allow valid inferences about health outcomes and access patterns across the state.

2.2 Statistical Analysis.

First, we calculated the distribution of the sample by sociodemographic and health-related characteristics. Second, we calculated the weighted prevalence of delayed or unmet medical care among California teenagers in 2023, stratified by each characteristic. Third, we fit a weighted logistic regression model to assess the association between covariates and the binary outcome of delayed or unmet medical care (yes/no). Covariates included age, self-reported gender, race/ethnicity, citizenship status, rural/urban residence, family size, difficulty getting by on family income, and health insurance status in the past 12 months. All analyses were conducted using R version 4.5.1, applying survey weights to generate population-representative estimates. We set the type I error rate at 5% for all statistical tests.

3. Results

3.1 Characteristics of Sample.

The modal age among participants was 14 years. Of 968 teenagers sampled, 52.0% were male and 48.0% were female. Most identified as Latino (41%) or non-Latino White (35%), followed by non-Latino Asian (13%) and non-Latino multiracial or other single-race groups (10.8%). A large majority were U.S.-born citizens (92%) and lived in urban areas (82%). Most teens lived in families of 2 to 3 members (51%). Most reported never having difficulty getting by on their family's income (61%). The majority of teens (98%) were insured for the entire past year.

3.2 Prevalence of Delayed or Medical Care by Characteristics.

Overall, the prevalence of delayed or unmet medical care was 6.3% among California teens (95% CI: 6.1%-6.5%; Table 2). By age, prevalence ranged from 2.7% (95% CI: 2.5%-3.0%) among 13-year-olds to 13.4% (95% CI: 12.7%-14.2%) among 16-year-olds. The prevalence was 4.2% for male teens (95% CI: 4.0%-4.5%) and 8.5% for female teens (95% CI: 8.1%-8.8%). By race and ethnicity, prevalence was highest among non-Latino teens identifying with other single races (16.7%; 95% CI: 14.9%-18.4%) or two or more races (15.2%; 95% CI: 13.8%-16.6%), compared to 8.0% for non-Latino White (95% CI: 7.5%-8.4%), 3.7% for Latino (95% CI: 3.5%-3.9%), and 2.7% for non-Latino Asian (95% CI: 2.5%-3.0%) adolescents. US-born citizens had a higher prevalence of 6.7% (95% CI: 6.4%-6.9%) than naturalized or non-citizen teens at 1.5% (95% CI: 1.4%-1.7%). The reported prevalence was slightly higher in urban areas, with 6.4% for adolescents in urban areas (95% CI: 6.2%-6.7%) and 5.0% for adolescents in rural areas (95% CI: 4.7%-5.3%). The prevalence was 5.2% (95% CI: 4.9%-5.5%) among teens from families of 2-3 members, 7.8% (95% CI: 7.4%-8.1%) among those from families of 4, and 6.6% (95% CI: 6.1%-7.2%) among those from families of 5 or more. The prevalence of unmet care increased with the level of financial strain, with prevalence rising from 6.4% (95% CI: 6.1%-6.7%) among those never reporting financial difficulty to 13.5% (95% CI: 12.8%-14.2%) among those reporting financial difficulty very often. By insurance status, 6.4% (95% CI: 6.1%-6.6%) of adolescents insured all year reported delayed or unmet care, compared with 4.5% (95% CI: 3.5%-5.4%) of those uninsured part year, and 0.0% (95% CI: 0.0%-0.0%) of those uninsured all year.

3.3 Regression Results.

The odds of delayed or unmet medical care increased with age (adjusted odds ratio [aOR]=1.28; 95% CI: 1.24-1.32). Compared to males, the odds were more than twice as high for females (aOR=2.15; 95% CI: 1.97-2.35). Relative to Latino adolescents, the odds were higher for non-Latino White (aOR=2.08; 95% CI: 1.91-2.26), non-Latino other single race (aOR=6.02; 95% CI: 5.07-7.14), and non-Latino adolescents of two or more races (aOR=6.39; 95% CI: 5.63-7.26), but not for non-Latino Asian adolescents (aOR=1.06; 95% CI: 0.95-1.18). Compared to naturalized or non-citizen adolescents, the odds were higher for US-born citizens (aOR=4.49;

95% CI: 3.94-5.10). Compared to rural residence, the odds were higher for adolescents living in urban areas (aOR=1.35; 95% CI: 1.25-1.46). Compared to teens in families of 2–3 members, the odds were higher for those in families of 4 members (aOR=1.54; 95% CI: 1.41-1.69) and 5 or more members (aOR=1.44; 95% CI: 1.27-1.64). Compared to adolescents who never reported financial difficulty, the odds were higher for those who reported their families “very often” had difficulty getting by (aOR=1.76; 95% CI: 1.58-1.95), but not for those reporting difficulty “not very often” (aOR=0.95; 95% CI: 0.87-1.04) or “somewhat often” (aOR=1.03; 95% CI: 0.91-1.16). Compared to adolescents insured all year, the odds were similar for those uninsured part year (aOR=1.15; 95% CI: 0.93-1.42).

Discussion

Our study reached three central findings regarding delayed or unmet medical care among adolescents in California. First, the overall prevalence of delayed or unmet was 6.3%, representing a substantial proportion of youth at risk of negative health outcomes. Second, disparities emerged across demographic subgroups, with higher prevalence among female adolescents, older teens, and those from racial/ethnic minority backgrounds. Third, socioeconomic vulnerabilities, particularly being uninsured at any point in the past 12 months and financial hardship, were strongly associated with increased odds of delayed or unmet medical care.

Our results align with national research documenting that adolescents face unique barriers to care, which place them at heightened risk of both acute and chronic health consequences^{4,25,26}. For instance, unmet medical care is associated with increased risk of infectious diseases, delayed diagnoses, and exacerbation of chronic conditions^{8,9,11,12}. Similarly, analyses of national survey data found that 26.4% of US adolescents in 2021 experienced some form of delayed or unmet care annually, a prevalence comparable to what we observed in California²².

California-specific studies have likewise documented the prevalence of unmet needs among adolescents, particularly those from low-income households. Previous research has shown that Medi-Cal–eligible teens and those in immigrant families are disproportionately affected by delayed access to preventive and acute care services²⁷. Consistent with our findings, these studies highlight that barriers such as cost, limited insurance coverage, and provider shortages contribute to gaps in timely care.

Additionally, our analysis extends prior literature by emphasizing risk factors related to poverty and insurance instability. We found that adolescents who were uninsured during any part of the year faced markedly higher odds of unmet medical care. Our results align with national findings that fragmented or unstable coverage exacerbates vulnerability^{28,29}. These results suggest that even short-term lapses in insurance can create long-lasting barriers to access, particularly when combined with structural barriers such as transportation difficulties and healthcare workforce shortages¹⁶.

These findings have important implications for California's healthcare system. Expansions of Medi-Cal eligibility and reductions in administrative barriers could help address gaps in healthcare coverage for adolescents at risk of being uninsured for part of the year. Policies that ensure continuous coverage through adolescence, simplify enrollment processes, and address provider shortages, particularly in rural and underserved urban areas, may reduce the prevalence of delayed or unmet care.

A key strength of this study is the use of state-level data, which provides a robust assessment of adolescent healthcare access in California and allows for nuanced analyses across sociodemographic groups. Additionally, the inclusion of multiple covariates enabled us to identify risk factors beyond basic demographic categories, offering insight into the structural determinants of unmet care. However, our study also had limitations. The reliance on self-reported measures may introduce recall or social desirability bias³⁰. Furthermore, the CHIS data did not distinguish between the types of medical care delayed or unmet, nor did it capture the length of delay, both of which could influence the severity of consequences.

In conclusion, delayed or unmet medical care among California adolescents remains a significant and inequitable challenge, with elevated risks among females, racial or ethnic minority youth, and those with unstable health insurance. Future research could build on this foundation by conducting longitudinal studies to examine the downstream health, developmental, and educational consequences of unmet medical care. By addressing structural and policy barriers, California has the opportunity to improve adolescent health outcomes and set a model for other states seeking to ensure equitable access to care.

Acknowledgements

I am deeply grateful to the many individuals who have contributed to the compilation of this study.

First and foremost, I would like to express my sincere gratitude to Dr. Samir Soneji for their guidance, expertise, and encouragement throughout the course of this journey. Their insights and advice have been instrumental in shaping the outcomes of this paper.

I also wish to extend my thanks to Polygence for providing the necessary resources and support, without which this project would have not have been possible.

Lastly, I would like to thank my family and friends for their helpful comments and persistent encouragement. Their support gave me the strength to persevere and has been a constant source of motivation.

Works Cited

- 1.Hargreaves DS, Elliott MN, Viner RM, Richmond TK, Schuster MA. Unmet Health Care Need in US Adolescents and Adult Health Outcomes. *Pediatrics*. 2015;136(3):513-520. doi:10.1542/peds.2015-0237
- 2.Alderman EM, Breuner CC, COMMITTEE ON ADOLESCENCE, et al. Unique Needs of the Adolescent. *Pediatrics*. 2019;144(6):e20193150. doi:10.1542/peds.2019-3150
- 3.Chang AR, Slopen N. Racial and Ethnic Disparities for Unmet Needs by Mental Health Condition: 2016 to 2021. *Pediatrics*. 2023;153(1):e2023062286. doi:10.1542/peds.2023-062286
- 4.Hargreaves DS, Elliott MN, Viner RM, Richmond TK, Schuster MA. Unmet Health Care Need in US Adolescents and Adult Health Outcomes. *Pediatrics*. 2015;136(3):513-520. doi:10.1542/peds.2015-0237
- 5.Girdwood TC, Silva SG, Maslow GR, Docherty SL. Disparities exist among US adolescents in the receipt of transition to adult healthcare services: the differential impact of social determinants of health, healthcare needs, and COVID-19. *Front Public Health*. 2024;12:1452418. doi:10.3389/fpubh.2024.1452418
- 6.Unmet need - Health, United States. June 26, 2023. Accessed September 6, 2025. <https://www.cdc.gov/nchs/hus/sources-definitions/unmet-need.htm>
- 7.Psaroudakis I, Quattrone F, Tavoschi L, et al. Engaging adolescents in developing health education interventions: a multidisciplinary pilot project. *Eur J Public Health*. 2020;30(4):712-714. doi:10.1093/eurpub/ckaa012
- 8.CDC. Flu Vaccine Effectiveness for Children and Older Adults. Flu Vaccines Work. October 11, 2024. Accessed August 31, 2025. <https://www.cdc.gov/flu-vaccines-work/risk-groups/index.html>
- 9.Flannery B, Reynolds SB, Blanton L, et al. Influenza Vaccine Effectiveness Against Pediatric Deaths: 2010–2014. *Pediatrics*. 2017;139(5):e20164244. doi:10.1542/peds.2016-4244
- 10.Fugl A, Andersen CL. Epstein-Barr virus and its association with disease - a review of relevance to general practice. *BMC Fam Pract*. 2019;20(1):62. doi:10.1186/s12875-019-0954-3
- 11.Torres SD, Kim CY, Das M, et al. Delays in Diagnosis and Treatment of Bacterial Meningitis in NYC: Retrospective Cohort Analysis. *The Neurohospitalist*. 2022;12(2):268-272. doi:10.1177/19418744211037319
- 12.CDC. Managing Chronic Health Conditions. Managing Health Conditions in School. December 11, 2024. Accessed August 31, 2025. <https://www.cdc.gov/school-health-conditions/chronic-conditions/index.html>
- 13.Gerall CD, DeFazio JR, Kahan AM, et al. Delayed presentation and sub-optimal outcomes of pediatric patients with acute appendicitis during the COVID-19 pandemic. *J Pediatr Surg*. 2021;56(5):905-910. doi:10.1016/j.jpedsurg.2020.10.008

14. Meng JF. Factors Associated With Not Receiving Mental Health Services Among Children With A Mental Disorder in Early Childhood in the United States, 2021–2022. *Prev Chronic Dis.* 2024;21. doi:10.5888/pcd21.240126
15. Educational opportunity and access to pediatric care are more intertwined than you might think. Brookings. Accessed August 31, 2025.
<https://www.brookings.edu/articles/educational-opportunity-and-access-to-pediatric-care-are-more-intertwined-than-you-might-think/>
16. Garney W, Wilson K, Ajayi KV, et al. Social-Ecological Barriers to Access to Healthcare for Adolescents: A Scoping Review. *Int J Environ Res Public Health.* 2021;18(8):4138. doi:10.3390/ijerph18084138
17. Olson LM, Tang S, Fong S, Newacheck PW. Children in the United States with discontinuous health insurance coverage. *N Engl J Med.* 2005;353(4):382-391. doi:10.1056/NEJMsa043878
18. Syed ST, Gerber BS, Sharp LK. Traveling Towards Disease: Transportation Barriers to Health Care Access. *J Community Health.* 2013;38(5):976-993. doi:10.1007/s10900-013-9681-1
19. Courtwright SE, Barr EA. Pediatric Nurse Practitioner Workforce Shortage Threatens Child Health Equity: Key Contributors and Recommendations. *J Am Assoc Nurse Pract.* 2023;35(11):661-665. doi:10.1097/JXX.0000000000000954
20. Wisk LE, Witt WP. Predictors of delayed or forgone needed health care for families with children. *Pediatrics.* 2012;130(6):1027-1037. doi:10.1542/peds.2012-0668
21. Anderson KE, McGinty EE, Presskreischer R, Barry CL. Reports of Forgone Medical Care Among US Adults During the Initial Phase of the COVID-19 Pandemic. *JAMA Netw Open.* 2021;4(1):e2034882. doi:10.1001/jamanetworkopen.2020.34882
22. Lebrun-Harris LA, Sappenfield OR, Warren MD. Missed and Delayed Preventive Health Care Visits Among US Children Due to the COVID-19 Pandemic. *Public Health Reports®.* 2022;137(2):336-343. doi:10.1177/00333549211061322
23. CHIS Design and Methods | UCLA Center for Health Policy Research. Accessed September 6, 2025.
<https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods>
24. Bureau UC. Data Profiles. Census.gov. Accessed September 6, 2025.
<https://www.census.gov/programs-surveys/acs/>
25. Bui AL, Ball AM, Ko LK, Ng M, Rivara FP, Coker TR. Foregone Preventive Care and Unmet Mental Healthcare Needs Among U.S. Youth. *Am J Prev Med.* 2025;69(2):107644. doi:10.1016/j.amepre.2025.107644
26. Elliott BA, Larson JT. Adolescents in mid-sized and rural communities: Foregone care, perceived barriers, and risk factors. *J Adolesc Health.* 2004;35(4):303-309. doi:10.1016/j.jadohealth.2003.09.015
27. Health Conditions and Health Care among California’s Undocumented Immigrants. Public Policy Institute of California. Accessed September 1, 2025.

- <https://www.ppic.org/publication/health-conditions-and-health-care-among-californias-un-documented-immigrants/>
28. Olson LM, Tang S, Fong S, Newacheck PW. Children in the United States with discontinuous health insurance coverage. *N Engl J Med*. 2005;353(4):382-391. doi:10.1056/NEJMsa043878
 29. DeVoe JE, Ray M, Krois L, Carlson MJ. Uncertain Health Insurance Coverage and Unmet Children's Health Care Needs. *Fam Med*. 2010;42(2):121-132.
 30. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc*. 2016;9:211-217. doi:10.2147/JMDH.S104807

RAPIDSS: Rat brain Automated Pipeline for DCE-MRI Segmentation and Signal-analysis

By Liana Kim

Abstract

Dynamic contrast-enhanced (DCE) Magnetic Resonance Imaging (MRI) datasets are complex, with numerous scans to review, making manual processing for analysis time-consuming and prone to error. We present a Python-based automatic processing pipeline that efficiently processes rat brain DCE-MRI scans. The pipeline automates Region of Interest (ROI) selection by skull stripping, performing atlas registration, and applying Gaussian filtering to reduce noise. Our pipeline extracts simple yet informative metrics, including time-to-peak (TTP), peak-to-baseline ratio, and peak enhancement ratio. These metrics allow visual comparisons of regional signal intensity dynamics through spreadsheets. Applied to rat datasets, our findings were efficiently and reliably produced, showing precise peaks and a clear pattern in the signal across the regions. Our automated pipeline reduces user variability when selecting ROIs and standardizes output. This open-source tool makes it a reliable tool for researchers investigating brain dynamics, such as the glymphatic system, using DCE-MRI.

Keywords

- DCE-MRI, automatic signal analysis, image processing, glymphatic system

Introduction

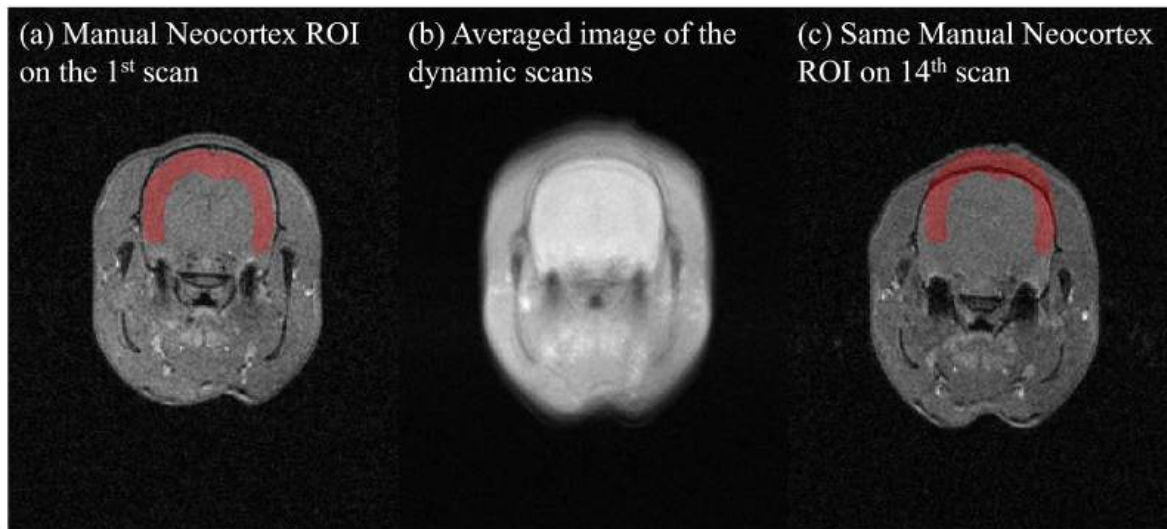


Figure 1: Manual ROI placement is misaligned relative to the baseline due to different motions of the head during the scan.

Dynamic contrast-enhanced MRI (DCE-MRI) is a method used to study the dynamic flow of a contrast agent in the brain by acquiring a time series of MRI images to track how

quickly and by how much the signal changes as the agent moves through tissue. Intravenous injections are a conventional DCE-MRI method used to track and study cerebral blood flow through the bloodstream. DCE-MRI with intrathecal contrast is used to investigate (a) the glymphatic system (brain clearance) via the intrathecal space surrounding the spinal cord and (b) the pharmaceutical kinetics via cerebrospinal fluid (CSF) circulation. It allows for a non-invasive readout of brain waste-clearance dynamics for direct analysis relevant to neurodegenerative diseases. Because the intrathecal DCE-MRI method is mainly limited to animal models, animal MRI plays a critical role in neuroscientific studies.

Specifically, rat brain DCE is commonly used because rats share physiological and genetic similarities with humans, making them suitable models for studying brain functions and diseases related to humans. Additionally, it enables invasive experiments that are nearly impossible to perform on humans. For example, experiments like intrathecal injections (injections into the space around the spinal cord) cannot be regularly done in humans due to ethical reasons. Due to these limitations, animal MRI offers a valuable tool for scientists to thoroughly analyze biological systems and processes *in vivo*, opening up new avenues for exploring the brain.

Animal (e.g, rat) DCE-MRI has become a critical tool in investigating the brain clearance process related to neurodegenerative diseases like Alzheimer's or Parkinson's disease. However, despite its importance, there is a lack of end-to-end, open-source DCE automatic MRI analysis tools to process the DCE-MRI more reliably and efficiently. For example, in manual ROI selection, the shape, size, and overall style of the ROI selected can vary based on each user's personal judgment or preference (Sohn et al., 2015). This inevitably leads to inaccuracies and poor reproducibility across time points and raters. However, through our processing pipeline's automation and standardization features, reproducibility improves. Consistent results regardless of user variability make DCE-MRI research more broadly accessible. In addition to user-driven ROI variability, issues such as motion inconsistency can affect reproducibility. As shown in Figure 1.b, image variation due to rat movement during dynamic acquisition is considerable. If this motion effect is to be reduced manually, all dynamic acquisitions must be checked, which can be a highly time-consuming process. Moreover, manual ROI selection on one scan can be shifted to a different region on another scan (Fig.1.c). This shifts the focus away from the desired area, resulting in inaccurate measurements of signal intensity.

While recent studies are focusing on new methods for skull stripping, atlas-based registration, and brain regional ROI selection, many of these approaches still rely on manual processes (Chwang et al., 2014), which are prone to human error and labor-intensive. Therefore, our study introduces an automated, open-source-based analysis pipeline that integrates skull stripping, atlas registration, and ROI extraction for rat brain DCE-MRI. Our open-access, automatic methods can precisely perform ROI selection and improve the accessibility of rodent MRI studies.

Methods

The DCE-MRI of six Sprague-Dawley rats was obtained from Seoul National University College of Medicine using a 3T animal MRI scanner. Before and after the intrathecal contrast injection, T1-weighted DCE-MRI scans were performed to acquire a time series of 26 consecutive scans over a 24-hour period, which included a pre-scan and the 0.5-hour scan, as shown in Figure 2.

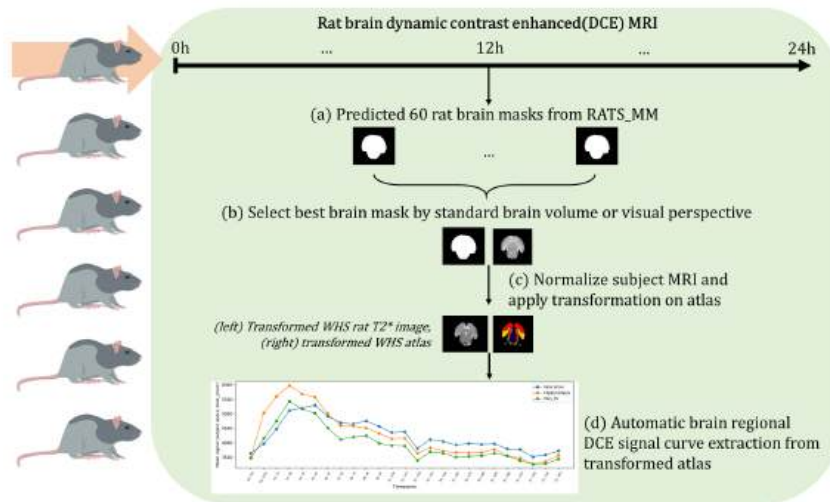


Figure 2: Schematic demonstration of the proposed rat brain automated pipeline for segmentation and signals of DCE-MRI.

Script / Module	Function (Stage)	Description of Processing Steps
<i>rats_autotune.sh</i> (Bash)	Skull stripping (Stage 1)	Grid-searches RATS parameters (threshold/volume/kernel), evaluates ~60 masks, selects volume $\approx 2000 \text{ mm}^3$; outputs <i>best_mask.nii.gz</i> and <i>best_brain.nii.gz</i> .
<i>whs_register_and_ext ract_flirt.sh</i> (Bash)	Atlas registration (Stage 1)	FLIRT 12-DOF affine aligns the subject $\leftrightarrow$ WHS atlas; applies the transform to the atlas labels; outputs the subject-space labels and QC overlay.
<i>autoROIselection_and_Figures_and_Signals.py</i> (Python)	ROI based time-series analysis (Stage 2)	Loads 4D DCE data, extracts regional mean signals using atlas ROIs, normalizes, computes DCE metrics (TTP, peak enhancement, peak-to-baseline), and generates plots and Excel/CSV files.

Table 1. Summary of the automated DCE-MRI analysis pipeline scripts and their functions of our open-access software (<https://github.com/lianakimn/AutoRATDCE-MRI>).

Skull stripping

Our study utilizes RATS_MM, an atlas-free brain extraction tool for rodent MRI (Oguz, Ipek et al., 2014). This program was applied to each scan to generate candidate brain masks across all time points. An internal algorithm is performed to test a range of pre-defined RATS_MM parameter combinations, including thresholds, volume cutoffs, and kernel sizes for each scan. Approximately 60 parameter combinations are produced for each scan (Figure 2.a, Table 1). Each mask generated by RATS_MM undergoes morphological operations such as erosion and dilation to smooth out the mask and retain its most significant connected component. This step removes non-brain tissue and noise outside the brain to produce a clean brain mask suitable for final evaluation and analysis.

From the 60 different parameter combinations for each time point, the optimal brain mask is automatically extracted based on its volume (Figure 2.b, Table 1). The code calculates a quantitative score across all masks to ensure that the mask volume is as close to the target rat brain reference volume of 2000 mm³. This score is calculated by finding the absolute difference between a mask's volume and 2000 mm³. A lower score indicated that the mask volume was closer to the reference. This is a good indication that the volume isn't too small that brain tissue was cut out, or too large that non-brain tissue was left in. This automatic selection minimizes manual tuning while preserving brain tissue.

Image normalization and Atlas registration

Once the optimal masks are chosen for each point, all brain images were first corrected for variations in signal intensity. This ensured that the brightness range matched the atlas input, making the MRI scans comparable across all subjects. Following intensity normalization, each brain image was registered to a standard rat brain atlas to account for the rats' movement during the 24-hour MRI scan.

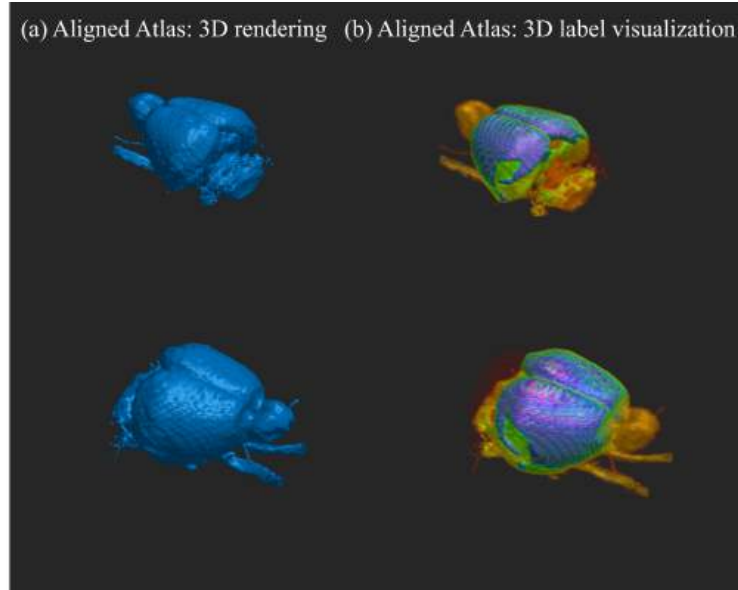


Figure 3: Atlas successfully reshaped to match the sample subject's (Subject #1) rat brain, aligned with its regions.

Using FSL FLIRT (Jenkinson, Mark et al., 2012), the Waxholm Space Atlas (WHS) for SD rat (Kleven, Heidi, et al., 2023) was aligned to each subject's 3D T1-weighted MRI volume (Figure 3.a). The actual subject's brain remained unchanged throughout atlas registration. The atlas was rotated, scaled, and stretched to match the subject's brain, ensuring it fit within the registration atlas's scale and orientation (Figure 3b, Table 1). We used 12-DOF affine registration, and atlas labels were resampled using nearest-neighbor interpolation to preserve integer values. This transformation process introduced geometric alignment for mapping labels, such as the cerebral cortex, neocortex, periventricular/periaqueductal regions, and hippocampus, to avoid variability in ROI extraction across scans (Figure 2.d, Table 1).

Image denoising

After atlas registration, the code performed Gaussian filtering to denoise each brain volume, reducing random noise without blurring anatomical structures, which stabilized the regional time–intensity curves..

ROI definition and signal extraction

Our ITK-SNAP label description file, WHS_SD_rat_atlas_v4_text.txt, assigns a unique index label (e.g., cerebellum = 5, subthalamic nucleus = 3) and RGB code to distinguish every brain region in the Waxholm atlas. Since each voxel in the 3D volume is assigned one of these index numbers, the text file acts as a “dictionary” of brain regions to translate voxel indices into human-readable brain regions. This makes segmentation easier to visualize, allowing for the immediate extraction of ROI for the desired brain region. Each subregion (e.g., the hippocampus, which is split into CA1, CA2, CA3, dentate gyrus, and subiculum) was automatically grouped by

function using `parseIdsByTerms`. By grouping multiple indices, the function allows researchers to flexibly define broader ROIs without being restricted to the atlas's detailed parcellation scheme.

In our study, we selected the neocortex, hippocampus, and the periventricular/periaqueductal (PV/PAG) regions for analysis. For each time point, the code averaged the DCE-MRI signal inside each ROI to make a time-intensity curve. The pre-contrast scan served as the baseline signal.

DCE metrics

From each ROI signal curve, the software computed several interpretable DCE metrics: time-to-peak (TTP), peak signal difference ratio (percent increase over baseline), and peak-to-baseline. Time-to-peak is defined as the time required to reach the highest intensity. The signal difference ratio is the ratio comparing the signal intensity at a time point to the baseline (pre) intensity. It highlights an increase or decrease in intensity relative to the starting point. The ratio is calculated by dividing the signal intensity at each time point by the signal intensity at the baseline. Peak-to-baseline is defined as the time required to return from its peak to the baseline signal level.

$$\text{Peak signal difference ratio} = \frac{\text{Peak signal} - \text{Baseline signal}}{\text{Baseline signal}} \times 100\%$$

Statistical differences between groups were tested using a two-sample t-test. The dynamic analysis and figure generation were written in Python using standard libraries (nibabel, numpy, pandas, and matplotlib). These were metrics used to calculate and analyze the dynamic brain signal changes throughout the 24-hour imaging session. To statistically compare the signal intensity of two groups at a given time point, a two-sample t-test was employed.

Results

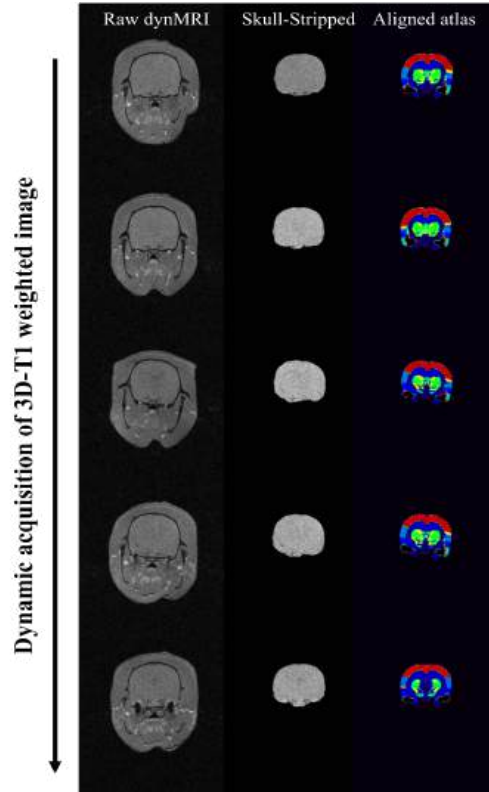


Figure 4: The last five dynamic scans of a sample subject from 0~24h data. The first column displays some variation in image position. However, after proper skull stripping and ROI segmentation through atlas registration, this motion effect is eliminated.

The automated, end-to-end pipeline was applied to control DCE-MRI data from six Sprague-Dawley rats, encompassing skull stripping, atlas registration, automatic ROI selection, and signal analysis. Across each time point, skull stripping precisely removed non-brain tissue, reducing noise. Then, it registered the Waxholm Space (WHS) onto the subject brain, in structural alignment with the specific brain regions of the extracted brain component (Figure 4). The dataset contains numerical labels for each brain region, which are turned into distinguishable colors. The different RGB colors serve as a visual confirmation that the segments were correctly assigned to correspond to the atlas definitions. Now, the same-colored atlas labels are consistent across all brain masks, and regions of interest can be extracted easily and reliably.

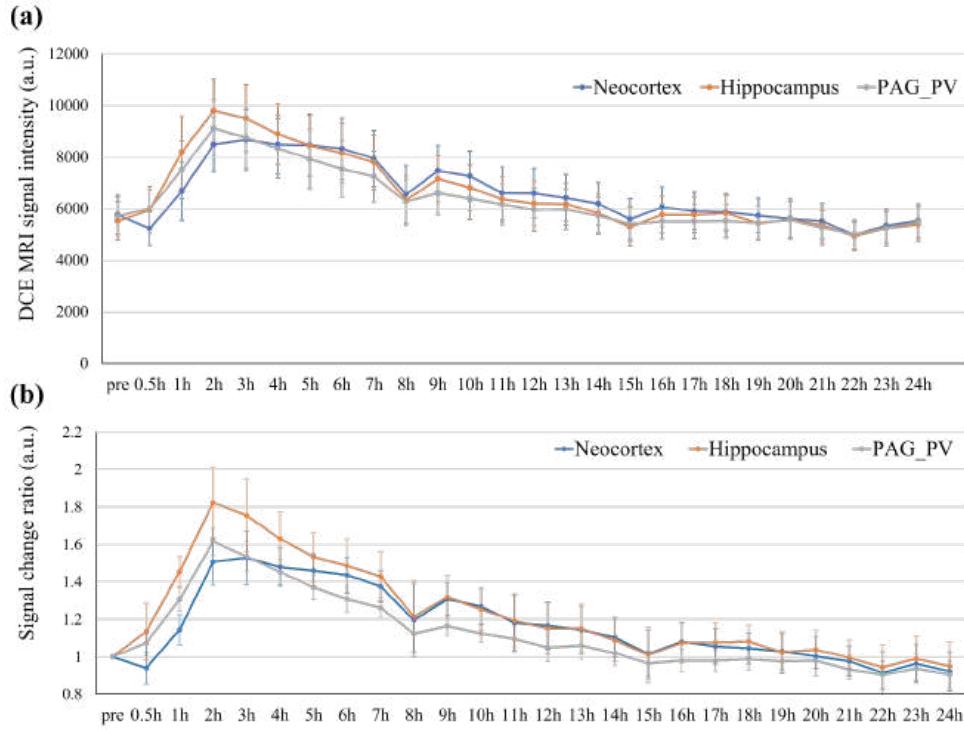


Figure 5: Signal change ratio and DCE-MRI signal intensities across all time points for Neocortex, Hippocampus, PAG_PV (average of data for all six rats)

Region	TTP (hours)	Peak Δ (%)	Peak-to-baseline (hours)
Neocortex	3.7 ± 1.6	+ 60.65%	17.3 ± 2.1
Hippocampus	2.0 ± 0.63	+ 84.53%	14.5 ± 2.4
Periventricular ROI	2.3 ± 0.52	+ 62.44%	10.4 ± 3.6

Table 2. Example DCE-MRI kinetic metrics extracted by the automated pipeline for three brain regions. (TTP = time to peak; Peak Δ = peak signal difference ratio; Peak-to-baseline.)

Following skull stripping and automatic ROI selection, the neocortex, hippocampus, and PAG/PV regions were distinguished. We extracted DCE-MRI signal intensity values and signal difference ratios across the three areas. The DCE-MRI signal intensity graph (Figure 5.a)

measures the raw signal intensity from the MRI scan at different time points. It shows how much the contrast agent accumulates and clears in the tissues of that specific region.

The hippocampus, neocortex, and PAG/PV follow similar trends over time. Within the first four hours, all three regions reach their highest peak. The hippocampus reached its highest DCE-MRI signal at approximately two hours (Table 2, Figure 5). The hippocampus also shows the highest percentage peak (~85%) signal difference ratio, compared to the neocortex (~61%) and periventricular ROI (~62%). Peak Δ indicates the most significant relative change in MRI signal intensity compared to baseline. The contrast agent in the neocortex brain region took the longest (~17) to reach baseline. The periventricular ROI, however, returned to baseline in ~10 hours. This indicates that the contrast agent left that region of the brain the quickest. Variability across regions differed, with the hippocampus showing the most consistent time-to-peak (2.0 ± 0.63 h), whereas the neocortex exhibited greater variability (3.7 ± 1.6 h). The greater standard deviation in neocortex TTP values may indicate variability even between rats, as they differ by subject. Depending on the subject, some show faster drainage in the cortex.

Compared to the manual ROI selection (Figure 1.c), automatic ROI selection (Figure 3, Figure 4) achieved accurate atlas alignment for reliable signal extraction.

Discussion

Our open-source, automated pipeline for processing rat brain DCE-MRI scans transformed the extracted signal intensities into regional curves and simple quantitative metrics, revealing clear trends across all regions. By automating each step, the pipeline reduces the risk of user variability caused by the inevitable movement of the rats over the 24-hour scan. The automated processes leading up to ROI signal extraction are efficient and reproducible. The pipeline was built with widely available resources—RATS_MM, FSL FLIRT, and standard Python scripts—making it convenient for researchers and laboratories to replicate easily.

As shown in Figure 5, the graph presents noticeable trends throughout the three regions of the subject's brain. The earlier time-to-peak (~2 hours) and higher peak (~9000 a.u.) in the hippocampus suggest a faster influx and greater capacity. Compared to the other brain regions, this faster influx may indicate that the glymphatic brain clearance system was highly active in this region. In the hippocampus, numerous toxic proteins are produced by metabolic processes (Bah et al., 2023). As a result, the hippocampal clearance system may be more critical than the neocortex or the periventricular ROI. However, as the clearance system in the hippocampus declines with age, waste products and toxic proteins may accumulate, leading to dementia (Jiang-Xie et al., 2025).

Our current approach contains limitations that need to be acknowledged. Our decision to use affine registration, for example, was a deliberate choice to prioritize speed and robustness in our dataset. However, images that exhibit severe distortion or extreme contrast may require an additional nonlinear registration step. For instance, if a rat happens to be abnormally long front-to-back or has a lesion such as a brain tumor, the brain is already deformed. Thus, the atlas

may not align well in terms of rotation or size. If the brain looks reasonably similar, the alignment will adjust to minor differences. Additionally, our study utilized DCE metrics (TTP, peak enhancement ratio, peak-to-baseline), which are less advanced than complete pharmacokinetic modeling, which we did not apply. These methods remain valuable, but were beyond the scope of the present work.

Future improvements could focus on improving the utility and strength of the pipeline. One improvement is an optional nonlinear registration module that would allow the atlas to be warped more flexibly to match any type of brain shape, thereby providing a more accurate ROI placement despite distortion. Although RATS_MM generally performs well, future work should focus on developing a plug-in deep-learning-based skull-stripper for exceptional cases, which would be beneficial in reducing preprocessing errors caused by unusual brain morphology or disease-related deformations. Additionally, conducting a broader validation study would be important when assessing not only spatial correspondence against expert-drawn labels but also for quantifying the efficiency using automation compared to manual methods. With standardized, open-source outputs, this tool will lower the barrier to entry for researchers and students in glymphatic-related DCE analysis research.

Conclusion

The efficiency and reproducibility of our automatic rat DCE-MRI analysis pipeline may be beneficial for studies that conduct detailed analysis of the glymphatic system. Our end-to-end system processing enables a comprehensive framework that reduces errors and ensures all data input flows smoothly, one step at a time. Unlike manual analysis, our pipeline eliminates the need for hands-on labor and significantly reduces variability in ROI selections across preclinical DCE studies. Our pipeline aims to aid in studies of the glymphatic system and neurodegenerative diseases through rodents, where dynamic CSF circulation and clearance are critical to understanding the underlying causes. The pipeline is reproducible across experiments, presenting accurate measurements of how quickly brain regions take up and clear a tracer.

Additionally, our code and documentation are publicly available on GitHub (MIT licensed) for adaptation and easy use. Such automated pipelines will improve the rigor of preclinical imaging research, just as open-source toolkits have done for human MRI research. Finally, through the standardization of analyses on a common atlas framework, our pipeline facilitates the easy comparison and aggregation of data, thereby accelerating discoveries regarding how factors such as aging, genes, or therapies impact brain physiology. We expect the community to build upon our work and add more advanced features to advance small animal MRI analysis.

Works Cited

- Bah, Thierno M., Dominic A. Siler, Aseel H. Ibrahim, Justin S. Cetas, and Nabil J. Alkayed. "Fluid Dynamics in Aging-Related Dementias." *Neurobiology of Disease*, vol. 177, 2023, p. 105986. ScienceDirect, <https://doi.org/10.1016/j.nbd.2022.105986>.
- Chwang, Wilson B et al. "Measurement of rat brain tumor kinetics using an intravascular MR contrast agent and DCE MRI nested model selection." *Journal of magnetic resonance imaging : JMRI* vol. 40,5 (2014): 1223-9. doi:10.1002/jmri.24469
- Jenkinson, Mark et al. "FSL." *NeuroImage* vol. 62,2 (2012): 782-90. doi:10.1016/j.neuroimage.2011.09.015
- Jiang-Xie, Li-Feng, Antoine Drieu, and Jonathan Kipnis. "Waste Clearance Shapes Aging Brain Health." *Neuron*, vol. 113, no. 1, 2025, pp. 71–81. ScienceDirect, <https://doi.org/10.1016/j.neuron.2024.09.017>.
- Kleven, Heidi, et al. "Waxholm Space Atlas of the Rat Brain: A 3D Atlas Supporting Data Analysis and Integration." *Nature Methods*, vol. 20, no. 11, Nov. 2023, pp. 1822–29, <https://doi.org/10.1038/s41592-023-02034-3>.
- Oguz, Ipek et al. "RATS: Rapid Automatic Tissue Segmentation in rodent brain MRI." *Journal of neuroscience methods* vol. 221 (2014): 175-82. doi:10.1016/j.jneumeth.2013.09.021
- Sohn, William S et al. "Influence of ROI selection on resting state functional connectivity: an individualized approach for resting state fMRI analysis." *Frontiers in neuroscience* vol. 9 280. 11 Aug. 2015, doi:10.3389/fnins.2015.00280

PARO Robot in Dementia Care: A Systematic Review With Narrative Synthesis

By Hillary Liu Sun

Abstract

Cognitive issues like dementia are common problems that those who are aging experience, often leading to behavioral and psychological symptoms (BPSD) and reduced quality of life (QoL). As technology continues to advance, there's an increasing interest in using socially assistive robots to help these patients. The most extensively studied robot is the seal-like, PARO robot. The aim of this study is to analyze evidence about the effectiveness of PARO in reducing BPSD and improving QoL. A systematic review with narrative synthesis was conducted, searching on PubMed for articles published between 2013 to 2025. Included studies were randomized controlled trials (RCTs) using PARO intervention to assess its effects on engagement, depression, and QoL in older adults with dementia. The articles were reviewed carefully and categorized into groups based on the specific outcomes studied. From an initial 71 articles, 44 full-text articles were assessed for their eligibility, with 8 ultimately included in the analysis. All 5 articles examining the effectiveness of PARO in improving engagement found that PARO had significant positive effects across visual, verbal, and physical engagement. 2 of the 3 studies examining PARO's effects on depression found that PARO reduced depressive symptoms, while 1 found that intervention had no effect. 2 of the 3 articles examining QoL found that PARO had significant positive effects, with one even finding that it was more effective than traditional animal-assisted therapy (AAT). Therefore, using PARO in dementia care can be effective in improving BPSD and is a promising alternative to AAT and pharmacological treatments.

Keywords

Dementia, PARO, Behavioral and Psychological Symptoms of Dementia (BPSD), Quality of Life (QoL), Engagement, and Depression

Introduction

By 2030, dementia is expected to affect up to 75.62 million people and 135.46 million people by 2050¹. While ongoing research to find a cure for dementia is vital, it is equally important to explore ways to alleviate the daily symptoms experienced by those living with the condition. As the disease progresses, around 97% of individuals with dementia are affected by behavioral and psychological symptoms, adding to the cost and burden of caring for them². BPSD constitute many mental and behavior disturbances, with apathy, depression, and agitation being the most common symptoms³. These symptoms often result in increased medication use, long-term hospitalization, and decreased QoL for the patients and caregiver^{4,5}.

Many individuals are routinely prescribed antipsychotics in order to manage severe agitation, aggression, and psychotic symptoms like hallucinations or delusions. However, past research has shown that many of these drug interventions are associated with negative side

effects that can increase mortality rate⁶. Nonpharmacological treatments have been proven to be safer and possibly have more benefits as they can be tailored to each individual's needs².

One type of nonpharmacological intervention is AAT, the use of animals in therapy sessions. AAT has been shown to calm agitated behavior and has positive effects on the quality of social interactions. This type of therapy has also been shown to decrease verbal aggression and reduce behavior problems. Despite these positive effects, AAT is not always possible since some senior healthcare settings and nursing homes don't accept animals as there are concerns with working with animals^{1,4}.

Therefore, robotic pet therapy/the use of socially assistive robots is being viewed as a viable substitute for AAT⁷. Although there are many different types of assistive robots that have been developed, the most extensively researched is the PARO robot, short for "personal robot"⁸. PARO is a seal-like robot that imitates animal behavior but also over time is able to develop its own "character" due to its artificial intelligence capabilities. It is in the shape of a baby harp seal with interactive features, able to respond to sound, touch, light, and being put out of balance by movement and vocalization^{4,9}. PARO was created to promote therapeutic results for those who interact with it; however, there have been a lot of inconsistent findings in PARO's effectiveness. Some studies claim that there were significant benefits to QoL and others claim there were no statistically significant differences from baseline to the end of the intervention. So, this study conducted a systematic review with narrative synthesis to evaluate the effectiveness of the PARO socially assistive robot in reducing BPSD and improving QoL, aiming to clarify ongoing controversy and provide strong, evidence-based insights.

Articles were identified using PubMed with search terms such as "PARO," "dementia," "BPSD," and other related terms. Studies were included if they were RCTs including older adults with dementia using PARO as an intervention, and measured changes in engagement, depression, or QoL. Each study was assessed for its quality using the Cochrane Risk of Bias (RoB 2) tool. After extracting key data, a narrative synthesis was conducted by categorizing all the studies based on the specific symptom targeted.



Figure 1: PARO robot

Methods

Articles were chosen for this paper through searches of publications on PubMed from 2013 to present day; though, some articles' full free texts were on other databases, such as Frontiers and Journal of the American Medical Directors Association (JAMDA). I used the search terms "socially assistive robots," "PARO," "elderly care," "neurocognitive disorder," "dementia," "behavioral and psychological symptoms of dementia (BPSD)," and "disruptive behaviors," combining them together as needed.

Inclusion and Exclusion Criteria

Included studies fulfilled the following criteria: the study design was a RCT; study participants were older adults with cognitive impairment, such as dementia or Alzheimer's; the intervention involved the PARO robot interacting with the participants; the researchers measured changes to agitation, depression, visual and physical engagement, or QoL; and the article had a full free text version.

Excluded studies fulfilled the following criteria: the study was written in any language other than English; research papers that were literature, meta-analysis, or systematic reviews; and articles that focused solely on the development of the PARO robot.

Data Extraction and Quality Assessment

For studies that met the inclusion criteria, I first assessed their quality using the Cochrane Risk of Bias (RoB 2) tool, version 2019¹⁰. Each study was evaluated based on its design and conduct, following the key areas identified in the RoB 2 tool, including bias arising from the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Then, the following information was extracted from each article: first author's name, publication year, country, participants' characteristics (age and sex), intervention with PARO (duration of exposure and number of times per week or month the intervention was used), control and intervention groups, and outcomes of engagement, depression, and QoL.

Narrative Synthesis

I reviewed the characteristics of each included study and categorized them into different groups based on the specific BPSD the article targeted, such as agitation, engagement, QoL, and depression. This categorization allowed me to identify certain patterns in the intervention focus and efficiently facilitate cross-study comparisons. The narrative synthesis focused on evaluating the effectiveness of PARO interventions by identifying commonalities in outcome trends across the studies, whether positive or negative. The aim of this analysis was to show symptom-specific variations in response to PARO, not only the overall effectiveness of the interventions. Using this approach, the study sought to identify which BPSD would be most responsive to PARO and to what extent were these responses consistent.

Results

Identification of Relevant Studies

The PubMed search results yielded 71 articles on the topic of socially assistive robots in dementia care. 62 articles were screened based on their titles and abstracts and 17 were excluded because they were either incomplete studies or non-RCT studies. In total, 44 articles were eligible for full-text screenings and 36 articles were excluded for the following reasons: they were not primary research articles, the full text was unavailable, there was no available data included, the research was focused on other social robots, the data was focused solely on the effects on sleep, and because the article was not in English. In the end, 8 articles were included in the study.

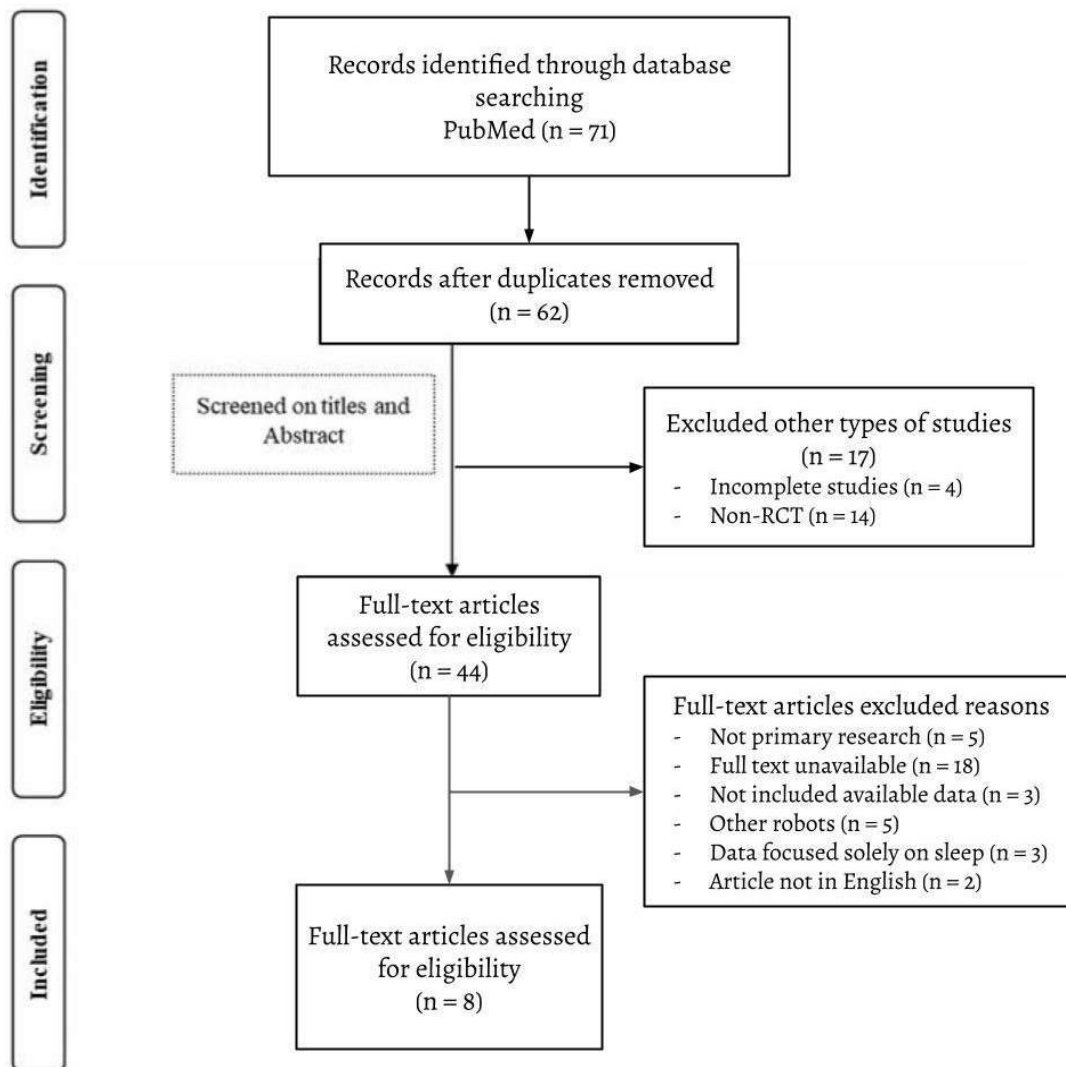


Figure 2: Flowchart of the study selection process

Characteristics and Quality of Included Studies

8 studies, from 2013 to 2025, included 1087 adults with dementia. The majority of participants with dementia were women with an average age of 84. These studies were

conducted in Japan, Australia, Denmark, America, New Zealand, and Spain. All of the studies were centered on the effectiveness of the PARO robot in reducing specific BPSD. Of the 8 articles, many of them studied the effect of intervention on multiple common symptoms of dementia. 5 articles reported on patient engagement, 3 on depression, and 3 on QoL. The majority of articles used the PARO robot in group settings; however, it was also sometimes utilized in individual contexts. Certain studies even intentionally let the interaction with PARO to be unstructured, allowing participants to engage spontaneously and freely, no matter the setting type. Therapy sessions mostly varied from 10 minutes to 1 hour sessions 2-3 times per week, with duration being around 2 or 3 months. A majority of the included RCTs showed a low or moderate risk of bias overall. Among the studies where the randomization process was applicable, all showed low risk bias. The assessment also showed that the studies demonstrated an overall low risk of bias in allocation concealment and incomplete outcome data. However, high risk of bias was consistently observed in the blinding of participants and personnel because the presence of the PARO robot made it immediately clear to the participants that they were part of the experimental group, therefore compromising the ability to blind them. Some areas, such as selective reporting and blinding of outcome assessment, were poorly reported, with almost half of them being rated as unclear.

	Randomization process	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting
Inoue et al. 2021 (Japan)	X	-	-	+	?	-
Jones et al. 2018 (Australia)	+	?	-	?	+	+
Liang et al. 2017 (New Zealand)	+	+	-	-	+	+
Moyle et al. 2017 (Australia)	+	+	-	+	+	?
Petersen et al. 2017 (U.S.)	+	+	-	?	?	+
Robinson et al. 2013 (New Zealand)	+	+	?	?	+	?
Soler et al. 2015 (Spain)	+	?	-	+	?	?
Thodberg et al. 2016 (Denmark)	+	+	-	?	+	+

Figure 3: Assessment of risk of bias; X mark: inapplicable, positive mark: low risk of bias, negative mark: high risk of bias, question mark: unclear bias.

Engagement

Study, Country	Participants	Intervention	Control	Measures	Findings
Jones et al. (2018)¹¹, Australia	138 people (age ≥ 60 , 73% female) at 9 facilities	15 min. Individual sessions 3x per week for 10 weeks	N/A	Cognitive impairment, agitation, engagement	- Agitation improved only in patients with low/moderate baseline agitation - Cognitive impairment improved from baseline - PARO sessions encouraged greater verbal and visual engagement
Liang et al. (2017)¹², New Zealand	91 rest home residents	1 hour sessions 2-3x/week for 6 weeks	Standard activities with caregivers (exercise, games, music)	Agitation, depression, engagement	- Participants in the PARO group showed more positive facial expressions and engaged in “conversations” with PARO, indicating improvements in agitation and depression - Participants in the PARO group interacted more with the staff and researchers than those in the control group - Although patients interacted with PARO, they were more engaged with live animals - PARO was suggested to be more helpful for individuals with “higher functioning capabilities” and at the “earlier stages of dementia”
Inoue et al. (2021)¹³, Japan	7 people (age ≥ 65 , 1 man)	At least 3x/week, 15-180 mins depending on participant response, over 1-3 months	N/A	Mood, engagement, feeling of caregiver burden	- 5/7 participants showed an improvement in mood and engagement - PARO allowed participants to interact with their family members by creating a sense of connection and belongingness - PARO use within the home context provided a sense of comfort by serving as a companion when facing unfamiliar individuals
Thodberg et al. (2016)⁹, Denmark	124 nursing home residents	10 minute sessions 2x/week for 6 weeks	1 group completed sessions with a dog, and another	Engagement	- Higher levels of cognitive impairment predicted more interaction with PARO and the animals

			group completed sessions with a toy cat		- High function levels meant longer conversations with visiting people - Engagement with the dog and PARO was higher than with the toy cat
Moyle et al. (2017)¹⁴, Australia	415 people (age ≥ 65)	15 minute sessions 3x/week for 10 weeks	Standard care	Engagement, agitation	- PARO group was significantly more engaged (visually and verbally), felt more pleasure (improved mood), and was slightly less agitated

Table 1: Studies focused on effect on engagement

5 studies examined the effectiveness of PARO on engagement and included 775 people with dementia. This was the most consistent benefit across all the studies. Engagement was measured through various indicators, such as visual, verbal, and physical interactions with PARO. 3 of the studies found that intervention with the PARO robot led to more visual engagement than standard care; one article even claimed that it was the most stimulated type of engagement (Jones et al. (2018)¹¹, Moyle et al. (2017)¹⁴, Thodberg et al. (2016)⁹). 3 other studies found that there was greater verbal engagement as well, saying that the pre and post intervention values were statistically different (Liang et al. (2017)¹², Moyle et al. (2017)¹⁴, Thodberg et al. (2016)⁹). Participants often commented on PARO's appearance and responded to PARO's noises based on their own interpretations of the noises to hold "conversations." And finally, 2 studies found that the PARO intervention facilitated physical touch and activity engagement because the ability of PARO to provide feedback triggered more engagement from the study participants (Inoue et al. (2021)¹³, Thodberg et al. (2016)⁹). While PARO was most effective in increasing engagement, some studies did note that the interest in engagement decreased during the intervention period (Liang et al. (2017)¹², Thodberg et al. (2016)⁹).

Depression

Study, Country	Participants	Intervention	Control	Measures	Findings
Robinson et al. (2013)¹⁴, New Zealand	40 people (age 55-100, 13 men)	1 hour sessions 2x/week for 12 weeks	Standard care	Loneliness, depression, quality of life	- Loneliness decreased significantly in the PARO group, suggesting that PARO is an adequate substitute for live animals in nursing home facilities - No significant effects found on depressive scores or QoL scores
Liang et al. (2017)¹², New Zealand	91 rest home residents	1 hour sessions 2-3x/week for 6 weeks	Standard activities with caregivers (exercise,	Agitation, depression	- Participants in the PARO group showed more positive facial expressions, indicating improvements in agitation and depression - Although patients interacted with PARO, they

			games, music)		were more engaged with live animals - Depressive symptoms showed improvements from baseline to 6 weeks; however, these symptoms then increased from 6-12 weeks
Petersen et al. (2017)⁴, United States	61 people (77% female, average age of 83.4 years)	20 minute sessions 3x/week for 3 months	Standard care	Mood, depression	- Rating for Anxiety in Dementia (RAID), Cornell Scale for Depression in Dementia (CSDD), pain medication, and behavior medication were significantly decreased in the treatment group - Researchers found that PARO intervention was a viable option for controlling symptoms of anxiety and depression in people with dementia

Table 2: Studies focused on effect on depression

3 studies examined the effectiveness of PARO on decreasing depression-like symptoms, as it is a very common symptom of dementia. In total, there are 192 people with dementia across these studies. 1 of the studies found that depression scores improved significantly in the treatment groups (Petersen et al. (2017)⁴). Another study initially found a slight decrease in depressive scores; however, after changing the pre-intervention depression scores, the decrease was negligible (Robinson et al. (2013)¹⁴). The same study also investigated loneliness, finding that there was a significant decrease in the scores from the PARO group compared to the baseline scores. They also clarified that loneliness is closely related and often impacts depression. The final article found that the PARO group didn't have any significant differences in depressive symptoms post-intervention compared to the control group that used standard care (Liang et al. (2017)¹²).

Quality of Life (QoL)

Study, Country	Participants	Intervention	Control	Measures	Findings
Robinson et al. (2013)¹⁴, New Zealand	40 people (age 55-100, 13 men)	1 hour sessions 2x/week for 12 weeks	Standard care	Loneliness, depression, quality of life	- Loneliness decreased significantly in the PARO group - Researchers concluded that improvements in loneliness may indicate improvements in other areas of life, as a person feels less socially isolated in their current living state - No significant effects found on depressive scores or QoL scores
Inoue et al.	7 people (age ≥ 65, 1	At least 3x/week,	N/A	Mood, engagement,	- 5/7 participants showed an improvement in mood, engagement, and QoL

(2021)¹³, Japan	man)	15-180 mins depending on participant response, over 1-3 months		quality of Life	- Most caregivers reported that participants displayed a positive reaction and behavior towards PARO and others
Soler et al. (2015)¹, Spain	Phase 1: 101 people (88% women, average age of 84.68 years) Phase 2: 110 people (90% women, average age of 84.7 years)	30-40 minute sessions 2x/week for 3 months	Standard care, NAO, DOG	Apathy, quality of life, cognitive impairment	Phase 1: Significant difference in Apathy scores between PARO group and CONTROL group, but not between PARO and NAO - Significant decreases were also reported in the Neuropsychiatric Inventory (NPI) scores, specifically irritability and total NPI scores Phase 2: Significant difference in QoL for PARO group: Higher than DOG and CONTROL group - Although no changes were observed in Phase 1, in the second phase, patients that worked with the conventional therapy showed improvement in QoL

Table 3: Studies focused on effect on QoL

3 studies examined the effects of PARO on patient QoL, with a total of 148 participants included. One article found that previous findings, which indicated that PARO improves quality of life when used in care facilities, were consistent when PARO was used at home (Inoue et al. (2021)¹³). While one study found that PARO had a more positive effect than traditional AAT with a dog, the QoL scores had no significant changes between the first or second phase (Soler et al. (2015)¹). Although the Quality of Life in Late-Stage Dementia (QUALID) scores did not improve, many other factors, such as reduced apathy, fewer night-time behavioral issues, and improvements in irritability and overall NPI scores, suggest that the intervention may still offer meaningful benefits to QoL, as these domains directly influence an individual's well-being. However, this study only found this result in 1 of its 2 phases. The final article found that QoL remained consistent between pre and post-intervention (Robinson et al. (2013)¹⁴).

Discussion

This systematic review with narrative synthesis study provided and evaluated comprehensive evidence on the effectiveness of the use of PARO for care of people with dementia, specifically for the relief of BPSD. The review identified 3 categories of specific outcomes commonly assessed across the literature: engagement, depression, and QoL. While outcomes for decreasing depression and improving QoL are less conclusive, the results support the use of PARO for enhancing engagement.

The findings show that the most consistent and notable benefit of using PARO was increasing engagement in those with dementia. All 5 of the studies reported an increase in at least one form of engagement, including visual/eye contact, verbal communication, and physical

interaction. Increasing any form of engagement is crucial to those with dementia because it can significantly improve their QoL and decrease possibilities of isolation, which can cause more cognitive impairment and behavioral problems¹⁶. PARO was able to increase engagement, specifically verbal engagement, because of its capabilities to provide interactive feedback. However, some studies noted there was a decline in engagement over time, asserting that AAT didn't have this decline. Yet, since AAT is not always feasible in long-term care facilities, PARO remains a viable and more accessible alternative. This research highlights the need for more research to be conducted to know how to sustain the interest of the people with dementia and stimulation during extended interventions with PARO.

In contrast, the findings for PARO's effect on decreasing depression in those with dementia were less consistent. Only one article found a significant reduction in depression scores post-intervention with PARO. Another article found there was a significant reduction in loneliness, something often closely related to depression, but found no significant change in the depression scores themselves. And the final article reported no statistically significant effect on the depressive scores. Although, they did observe positive facial expressions during the interventions. These clashing findings may be the result of different measurement tools used/the lack of a standardized depression scale. Additionally, the differences in the type of intervention may have influenced the outcomes.

For QoL, the findings leaned towards a positive effect after PARO intervention. 2 out of the 3 studies found improvements in QoL. One article even stated that the PARO intervention had not only greater improvements than a different social robot (NAO) and AAT but also PARO specific benefits, offering the idea that PARO is uniquely effective in improving QoL for people living with dementia. However, one of the articles did not find a statistically significant change in QoL scores. They only found a slight reduction in loneliness. This discrepancy in results may be because some articles used a broader QoL assessment rather than the QUALID, dementia specific, QoL assessment¹⁷.

Future research should focus on exploring different strategies to sustain engagement with PARO over extended periods, such as personalized or rotating stimuli. Additionally, this research would benefit from the use of standardized dementia-specific measurement tools for depression and QoL, allowing for consistent comparisons across multiple studies. Larger and more diverse samples of participants could also further strengthen the consistency of future findings, leading to more robust evidence for PARO's effectiveness. To ensure that PARO is used correctly and appropriately, it may also be necessary to train the staff in how to use PARO in group and one-on-one therapy sessions.

Limitations

There are some limitations to this study. The first is the sample size. Because I didn't have access to all of the research databases, many of the times, I was unable to access full-text articles. Therefore I was limited to the articles I could include in this study. The second is heterogeneity. Because I didn't have access to a wide range of articles, many of the articles

included in this study employed different tests to create a baseline and test for the effects in the post-intervention assessment. This made it more difficult to pool data together as some results may be because of the usage of different standardized assessments.

Conclusion

Overall, this review shows that using PARO therapy in dementia care can be effective and has potential, especially in decreasing BPSD. The studies showed an overwhelming positive impact on increasing patient engagement. While the results on improving depression and QoL have inconsistencies, the overall outcome was still positive. As a nonpharmacological intervention, PARO is a promising alternative to traditional treatments, where they aren't feasible.

Works Cited

- 1.Valenti Soler, Meritxell, et al. "Social Robots in Advanced Dementia." *Frontiers in Aging Neuroscience*, vol. 7, 3 Sept. 2015, <https://doi.org/10.3389/fnagi.2015.00133>.
- 2.Kales, H. C., et al. "Assessment and Management of Behavioral and Psychological Symptoms of Dementia." *BMJ*, vol. 350, 2 Mar. 2015, <https://doi.org/10.1136/bmj.h369>.
- 3.Lu, Li-Chin, et al. "Effectiveness of Companion Robot Care for Dementia: A Systematic Review and Meta-Analysis." *Innovation in Aging*, vol. 5, no. 2, 24 Apr. 2021, <https://doi.org/10.1093/geroni/igab013>.
- 4.Petersen, Sandra, et al. "The Utilization of Robotic Pets in Dementia Care." *Journal of Alzheimer's Disease*, vol. 55, no. 2, 19 Nov. 2016, pp. 569–574, content.iospress.com/download/journal-of-alzheimers-disease/jad160703?id=journal-of-alzheimers-disease%2Fjad160703, <https://doi.org/10.3233/jad-160703>.
- 5.Mordoch, Elaine, et al. "Use of Social Commitment Robots in the Care of Elderly People with Dementia: A Literature Review." *Maturitas*, vol. 74, no. 1, Jan. 2013, pp. 14–20, www.sciencedirect.com/science/article/pii/S0378512212003489, <https://doi.org/10.1016/j.maturitas.2012.10.015>. Accessed 20 Sept. 2025.
- 6.Gill, Sudeep S, et al. "Antipsychotic Drug Use and Mortality in Older Adults with Dementia." *Annals of Internal Medicine*, vol. 146, no. 11, 2007, pp. 775–86, www.ncbi.nlm.nih.gov/pubmed/17548409, <https://doi.org/10.7326/0003-4819-146-11-200706050-00006>. Accessed 20 Sept. 2025.
- 7.Abbott, Rebecca, et al. "How Do "Robopets" Impact the Health and Well-Being of Residents in Care Homes? A Systematic Review of Qualitative and Quantitative Evidence." *International Journal of Older People Nursing*, vol. 14, no. 3, 9 May 2019, <https://doi.org/10.1111/opn.12239>. Accessed 20 Sept. 2025.
- 8.Yu, Clare, et al. "Socially Assistive Robots for People with Dementia: Systematic Review and Meta-Analysis of Feasibility, Acceptability and the Effect on Cognition, Neuropsychiatric Symptoms and Quality of Life." *Ageing Research Reviews*, vol. 78, June 2022, p. 101633, <https://doi.org/10.1016/j.arr.2022.101633>. Accessed 20 Sept. 2025.
- 9.Thodberg, Karen, et al. "Behavioral Responses of Nursing Home Residents to Visits from a Person with a Dog,a Robot Seal or AToy Cat." *Anthrozoös*, vol. 29, no. 1, 2 Jan. 2016, pp. 107–121, <https://doi.org/10.1080/08927936.2015.1089011>.
- 10.Higgins, Julian PT. "Revised Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2)." *CoChrane Methods Bias*, 22 Aug. 2019, drive.google.com/file/d/19R9savfPdCHC8XLz2iiMvL_71IPJERWK/view. Accessed 20 Sept. 2025.
- 11.Jones C, Moyle W, Murfield J, Draper B, Shum D, Beattie E, Thalib L. Does Cognitive Impairment and Agitation in Dementia Influence Intervention Effectiveness? Findings From a Cluster-Randomized-Controlled Trial With the Therapeutic Robot, PARO. *J Am Med Dir Assoc*. 2018 Jul;19(7):623-626. doi: 10.1016/j.jamda.2018.02.014. Epub 2018 Apr 13. PMID: 29656838.

12. Liang, Amy, et al. "A Pilot Randomized Trial of a Companion Robot for People with Dementia Living in the Community." *Journal of the American Medical Directors Association*, vol. 18, no. 10, Oct. 2017, pp. 871–878, <https://doi.org/10.1016/j.jamda.2017.05.019>. Accessed 20 Sept. 2025.
13. Inoue, Kaoru, et al. "Exploring the Applicability of the Robotic Seal PARO to Support Caring for Older Persons with Dementia within the Home Context." *Palliative Care and Social Practice*, vol. 15, Jan. 2021, p. 263235242110302, <https://doi.org/10.1177/26323524211030285>. Accessed 20 Sept. 2025.
14. Moyle, Wendy, et al. "Use of a Robotic Seal as a Therapeutic Tool to Improve Dementia Symptoms: A Cluster-Randomized Controlled Trial." *Journal of the American Medical Directors Association*, vol. 18, no. 9, Sept. 2017, pp. 766–773, <https://doi.org/10.1016/j.jamda.2017.03.018>. Accessed 20 Sept. 2025.
15. Robinson, Hayley, et al. "The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial." *Journal of the American Medical Directors Association*, vol. 14, no. 9, 1 Sept. 2013, pp. 661–667, www.sciencedirect.com/science/article/pii/S1525861013000972?casa_token=WptUtGE5LSgAAAAA:xLw_9pk58FT7CoCsZ92SuAs4xZxZ4aPm-Mqd7GvOiDlvow23NJwZ0onY0ed99mBG9pyDrigY-Q, <https://doi.org/10.1016/j.jamda.2013.02.007>. Accessed 27 Oct. 2021.
16. Guarnera, Jade, et al. "The Impact of Loneliness and Social Isolation on Cognitive Aging: A Narrative Review." *Journal of Alzheimer's Disease Reports*, vol. 7, no. 1, 29 June 2023, pp. 699–714, pmc.ncbi.nlm.nih.gov/articles/PMC10357115/, <https://doi.org/10.3233/ADR-230011>. Accessed 20 Sept. 2025.
17. Tsang, Winnie, et al. "Quality of Life Measurement Tools for People with Dementia and Intellectual Disabilities: A Systematic Review." *Journal of Applied Research in Intellectual Disabilities*, vol. 36, no. 1, 15 Nov. 2022, <https://doi.org/10.1111/jar.13050>.

Feminism: Styled, Scripted, and Syndicated By Tanisha Chhabra

Introduction

We often think television reflects society—but what if it also rewrites our movements, making rebellion look safe, pretty, and easy? From the boardrooms of 1980s sitcoms to the confessionals of reality TV, American media has turned female ambition into primetime content, often flattening feminist ideals into digestible narratives of personal success. As Deborah Paredez points out in *Selenidad*, cultural memory is not simply preserved—it is staged, performed, and often reshaped to fit what is palatable for mass audiences. In the same way that Selena's legacy was adapted into a symbol of success while difficult realities about race and gender were softened or erased, feminist ideals were also reshaped by popular media to appear less radical. Rather than dismantling gender hierarchies, television portrayals often reaffirm them—suggesting that women can succeed, so long as they remain likable, attractive, and non-threatening. In doing so, television has redefined feminism for the masses: not as collective struggle, but as individual achievement wrapped in marketable aesthetics.

The Emergence of the Career Woman in 1980s Television

Far from revolutionizing gender roles, 1980s television negotiated a fragile compromise between feminism and patriarchy—one that celebrated women's ambition only when it served marketable, non-threatening narratives. Programs such as *The Mary Tyler Moore Show*, *One Day at a Time*, and *Alice* emerged as significant sites for what critics identified as a "feminist" subject position in American broadcast television (Rabinovitz). These shows were frequently cited by publications aimed at female audiences, including *Ms. Magazine* and *Working Woman*, as exemplary feminist representations. "Mary" was often positioned by popular critics as symbolizing the independent woman of the 1970s, establishing a template that would influence female characterization in the following decades (Rabinovitz), but their portrayal of "independent women" subtly reinforced traditional hierarchies. For example, while Mary's protagonist worked in a newsroom, her storylines often centered on romantic fulfillment, implying professional success was incomplete without domestic approval. The sitcom format, in particular, functioned to renegotiate sociopolitical differences as cultural affirmations of patriarchal values (Rabinovitz). This pattern continued through the 1980s, where shows featuring career women frequently counterbalanced their protagonists' professional ambitions with storylines emphasizing traditional feminine qualities and relationships (Andrews and McNamara). The tension between progressive characterization and conventional gender expectations defined this era's portrayal of ambitious women. By the mid-1980s, this pattern crystallized into a formula: female characters could excel in careers, but only if they remained palatably feminine—nurturing, emotionally available, and ultimately deferential to male authority (Moritz). This established a narrow definition of "acceptable feminism" in the cultural imagination – one that acknowledged women's professional capabilities while insisting they not challenge fundamental gender hierarchies.

Sitcoms and the Domestication of Feminist Ideals

This domestication of feminism didn't disappear in the 1990s; it evolved, with sitcoms weaponizing humor to frame gender equality as absurd or unnatural. While some sitcoms introduced feminist elements, many popular shows reinforced traditional gender stereotypes. What changed, however, was the hierarchical relationship between genders. While sitcoms modified the roles of women in an effort to keep pace with changing social norms, they frequently failed to meaningfully alter traditional masculine narratives (Andrews and McNamara). This resulted in a superficial role reversal: women became the capable, responsible figures while men were increasingly portrayed as childlike or incompetent – a dynamic that created comedy without genuinely challenging gender hierarchies (Rabinovitz). This was the rise of the "fun-killer" trope juxtaposing female competence with male buffoonery, implying that gender equality would destroy familial joy (Andrews and McNamara). Even seemingly progressive shows like *Murphy Brown* undermined their protagonist's independence by framing her careerism as a trade-off for personal happiness (Burger). This narrative equated feminism with individual sacrifice allowing audiences to applaud "strong women" while resisting structural change (Rabinovitz).

The Reality TV Revolution and Celebrity Feminism

Reality TV would later inherit this distorted framing, repackaging feminism as a personal brand rather than a political movement. The rise of reality television in the late 1990s and early 2000s introduced new dynamics in the portrayal of ambitious women. Unlike scripted programming, reality shows purported to present "authentic" female experiences, including women's professional ambitions and interpersonal relationships (Keren). This format created new opportunities for representation but also established problematic frameworks for understanding feminism and female empowerment. Reality television frequently positioned women's professional success within a neoliberal framework that emphasized individual achievement over structural critique. Success was framed as the result of personal determination rather than the dismantling of systemic barriers, reinforcing individualistic narratives that aligned with market values rather than collective social change (Keren). Marwick's analysis of *The Swan* also illustrates how reality makeover shows frame cosmetic surgery as a moral imperative, promoting a hyper-stylized version of femininity. This portrayal reinforces the idea that women's self-worth is tied to physical appearance and compliance with traditional beauty norms (Marwick). Shows centered on female celebrities or entrepreneurs presented ambition as praiseworthy only when it aligned with hyper-feminine aesthetics. This created a version of feminism palatable to mainstream audiences because it rarely challenged power structures. Celebrity media coverage has also significantly shaped public perceptions of feminism, often by presenting it through the lens of individual success stories. For instance, Beyoncé's 2014 MTV Video Music Awards performance, where she stood before a large "FEMINIST" sign, brought feminist discourse into mainstream pop culture, making it more accessible to a broad audience. This act was praised for challenging stereotypes and redefining feminism as a movement for equality (Bennett).

However, some scholars argue that such endorsements can oversimplify feminist ideals. Celebrity feminism can trivialize women's rights and reduce feminism to a publicity tool rather than a movement for systemic change. Celebrities could advocate for equal pay or reproductive rights while simultaneously reinforcing conventional beauty standards and consumer capitalism – a contradiction that lost the true meaning of feminism, prioritizing market-friendly individualism over substantive challenges to power structures (Christensen).

Professional Independence vs. Domestic Fulfillment

Nowhere was this individualism more damaging than in TV's persistent framing of professional success as incompatible with personal fulfillment. Female characters pursuing careers were frequently depicted struggling with romantic relationships, family obligations, or personal happiness – a narrative pattern that implicitly questioned whether women could truly "have it all". While male characters were rarely asked to choose between career and family (Andrews and McNamara). Shows like *Ally McBeal* in the 1990s positioned their protagonists as successful in their careers but unfulfilled in their personal lives, suggesting that professional women inevitably sacrificed domestic happiness (Burger). Such portrayals contributed to mainstream perceptions that feminism demanded sacrifices rather than seeking to restructure society to accommodate multiple aspects of fulfillment (Andrews and McNamara). Even shows like *Police Women of Broward County*, which showcased women in male-dominated fields, undercut their subjects by reducing them to working moms first, professionals second (Cox). This framing implicitly suggests that no matter their career achievements, women's primary responsibilities remain in the home. Work-life imbalance was framed as personal failure with television rarely depicting structural solutions to work-life conflicts such as affordable childcare, equitable division of domestic labor, or workplace policies supporting family responsibilities (Keren). Instead, work-life balance was portrayed as an individual challenge requiring personal sacrifice rather than societal change.

The Double Bind of the Ambitious Woman

Moreover, throughout the decades since the 1980s, television has frequently placed ambitious female characters in a double bind: criticized for being too career-focused and insufficiently feminine, or dismissed as unserious when exhibiting traditionally feminine traits. This contradictory expectation reinforced the impossible standards for women, regardless of their approach to balancing career and personal identity (Andrews and McNamara). The 'ice queen' and the 'ditzy professional' tropes made female ambition seem unnatural. The "ice queen" executive prioritizing career over personal relationships contrasted with the "ditzy" professional whose femininity undermined her credibility (Rabinovitz). Both character types suggested that women's professional ambitions were somehow unnatural or required significant personal compromise. Few shows presented ambitious women achieving success while maintaining authentic personal identities that transcended simplistic gender expectations (Andrews and McNamara). These limiting tropes contributed to mainstream perceptions that feminism

demanding women choose between professional achievement and authentic self-expression undermining the movement's goal of expanding choices for all individuals (Christensen). By framing ambitious women as exceptional figures navigating extraordinary challenges, television reinforced the notion that professional success remained fundamentally unnatural for women rather than a normal expression of human capability.

Beyond the Screen

These portrayals however don't just stay on-screen. Research from the Geena Davis Institute shows that media images strongly influence young women's career aspirations: when female characters are shown primarily in supporting or beauty-centered roles, girls are less likely to see themselves as future leaders (Mohammed). Martha Lauzen's *Boxed In* study also found that limited portrayals correlate with lower career confidence among women, particularly in leadership and STEM fields. By rarely depicting structural support for women—like childcare, equal pay, or workplace flexibility—television reinforces the idea that women must navigate work-life conflicts alone. Furthermore, media scholars Lauzen and Dozier found that when women held executive production roles, they hired more female writers and created more powerful female characters. Their 1999 study showed that female executive producers significantly influenced character development, giving women more agency, complexity, and dialogue that challenged traditional gender norms. This research reveals how representation behind the scenes can counterbalance the domestication of feminism often seen in sitcoms.

Solution

So to authentically represent feminist ideals, it's imperative to integrate feminist frameworks into the essence of television production. According to research by Dr. Lauzen in 2023-24, the percentage of women working as creators on original television series appearing on broadcast networks and streaming services declined for the second year in a row from just 30% of creators in 2021-22, down to 23% last year. Unsurprisingly, the percentage of female characters in speaking and major roles also declined. A feminist framework would necessitate a comprehensive reevaluation of the industry's structural dynamics, including ownership patterns, labor practices, and decision-making processes. Byerly and Ross introduce a feminist political economy perspective that critically examines how gendered power relations influence media production, distribution, and consumption, highlighting the systemic barriers that women face in the media industry. They argue that without addressing these foundational issues, efforts to improve on-screen representation risk being superficial and unsustainable. Furthermore, Zhang and Harvey advocate for restructuring labor relations by prioritizing pay equity, leadership access, and workplace support for marginalized groups. Rather than merely diversifying narratives, this approach seeks to transform the material and ideological foundations of media production, ensuring that feminist ideals are embedded not just in stories told but in the very processes of their creation. Addressing gender disparities in television production also requires targeted interventions such as mentorship programs and equitable hiring practices. Mentorship

has been identified as a critical factor in supporting women's career advancement in media by providing women with the necessary guidance, networks, and opportunities to navigate and succeed in a male-dominated industry (Byerly and Ross). Such programs can demystify the pathways to leadership roles and empower women to pursue and attain positions of influence. Equitable hiring practices are equally essential in fostering a diverse and inclusive workforce. Milner and Gregory, argue that without intentional efforts to dismantle systemic patterns that preserve disparities —through transparent recruitment processes, unbiased evaluation criteria, and accountability measures—gender inequalities will persist. Implementing equitable hiring practices ensures that women have fair access to opportunities and that their contributions are valued and recognized. These structural changes are vital for creating a media landscape that genuinely reflects and promotes feminist values, leading to more authentic and diverse storytelling. However in proposing this solution we must recognise meaningful and lasting structural change within a complex industry like television is time-consuming and requires sustained effort. Additionally, production companies and key decision-makers often prioritize profit margins over ideological shifts. So unless structural reforms demonstrably lead to increased revenue or reduced costs, they may be seen as secondary to financial objectives and therefore deprioritized.

Conclusion

For over four decades, American television has functioned as feminism's most seductive saboteur—applauding women's individual achievements while erasing the collective struggle required to make them possible. The trajectory from *The Mary Tyler Moore Show*'s charming independence to *The Bold Type*'s corporate "activism" reveals an unsettling pattern: each wave of feminist progress onscreen has been deliberately stripped of political teeth, repackaged as personal triumph rather than systemic overhaul (Andrews and McNamara). American television and the media industry exerted significant influence on mainstream perceptions of feminism—and weaponized its imagery to stall progress. The result has been a cultural understanding of feminism that celebrates women's professional accomplishments while rarely questioning the underlying structures that make those accomplishments more difficult for women than men (Rabinovitz). Yet the 2020s offer a subversive opportunity. Streaming platforms' appetite for complex narratives proves audiences crave stories that refuse to equate feminism with likability (Andrews and McNamara). As Lauzen and Dozier argue, real progress in television portrayals of women requires structural change within the industry itself. Increasing the number of women in leadership roles behind the camera is essential to ensuring feminist ideals are not only represented but authentically explored. By embedding feminist frameworks into the political economy of television production and instituting mentorship and equitable hiring practices, the industry can move beyond tokenistic representations.

Works Cited

- Andrews, Maggie, and Sallie McNamara, editors. *Women and the Media*. Routledge, Apr. 2014, <https://doi.org/10.4324/9780203074121>.
- Bennett, Jessica. "How to Reclaim the F-Word? Just Call Beyoncé." *TIME*, Time, 26 Aug. 2014, time.com/3181644/beyonce-reclaim-feminism-pop-star/.
- Burger, Alissa. "The 1990s Teen Horror Cycle: Final Girls and a New Hollywood Formula." *Journal of Popular Film and Television*, 2018, <https://doi.org/10.1080/01956051.2019.1608787>.
- Byerly, Carolyn, and Karen Ross. "Women and Production: Gender and the Political Economy of Media Industries." *Women and Media*, Wiley, Jan. 2006, pp. 75–95, <https://doi.org/10.1002/9780470774908.ch5>.
- Christensen, Kyle. "'Look What You Did to Me!': (Anti)Feminism and Extratextuality in the Remake of a Nightmare on Elm Street (2010)." *Journal of Film and Video*, vol. 68, no. 2, University of Illinois Press, Apr. 2016, pp. 29–45, <https://doi.org/10.5406/jfilmvideo.68.2.0029>.
- Cox, Nichole. "Communicating Gender on Reality TV - National Communication Association." National Communication Association -, 19 July 2016, www.natcom.org/publications-library/communicating-gender-reality-tv/.
- Keren, Hila. "Women in the Shark Tank: Entrepreneurship and Feminism in a Neoliberal Age." *Columbia Journal of Gender and Law*, vol. 34, no. 1, Columbia University Libraries, Dec. 2016, pp. 75–123, <https://doi.org/10.7916/d81g0zth>.
- Lauzen, Martha. "Women on Screen and behind the Scenes on Broadcast and Streaming Television in 2023-24." 2024, womenintvfilm.sdsu.edu/wp-content/uploads/2024/08/2023-24-Boxed-In-Report.pdf
- Marwick, Alice. "There's a Beautiful Girl under All of This: Performing Hegemonic Femininity in Reality Television." *Critical Studies in Media Communication*, 2025, <https://doi.org/10.1080/15295030903583515>.
- Milner, Susan, and Abigail Gregory. "Time for a Change: Women, Work, and Gender Equality in TV Production." *Media, Culture & Society*, vol. 44, no. 2, SAGE Publications, Oct. 2021, pp. 286–302, <https://doi.org/10.1177/016344372111045525>.
- Mohammed, Amina J. "Looking Back, Moving Forward." *Geena Davis Institute*, vol. 52, no. 2, Dec. 2015, pp. 30–32, <https://doi.org/10.18356/affb6a6d-en>.
- Moritz, Marguerite J. "American Television Discovers Gay Women: The Changing Context of Programming Decisions at the Networks." *Journal of Communication Inquiry*, vol. 13, no. 2, SAGE Publications, July 1989, pp. 62–78, <https://doi.org/10.1177/019685998901300208>.
- Paredes, Deborah. "Selenidad: Selena, Latinos, and the Performance of Memory." *Latino Studies Review*, vol. 4, no. 1, 2012, pp. 45–58.
- Rabinovitz, Lauren. "Sitcoms and Single Moms: Representations of Feminism on American TV." *Cinema Journal*, vol. 29, no. 1, JSTOR, 1989, p. 3, <https://doi.org/10.2307/1225298>.
- Zhang, Tracy Ying, and Alison Harvey. "Introduction: Feminist Media Production and Beyond." *Global Media and China*, vol. 9, no. 2, SAGE Publications, Apr. 2024, pp. 135–39, <https://doi.org/10.1177/20594364241247671>.

Correlation Between the Common Psychological Responses Athletes Experience Post-injury By Yeonjae Kim

Abstract

Sports injuries are a common aspect of an athlete's career, but often result in both physical and psychological setbacks that greatly affect one's career. While previous research has highlighted the mental health challenges faced by athletes post-injury, there has been limited research that has analyzed the interconnectedness of their emotional responses during the recovery phase. This paper aims to investigate the psychological responses athletes experience after an injury, specifically focusing on emotions such as denial, grief, fear, and optimism. Through a literature review of research from the past three decades, this study provides an overview of common emotional responses athletes feel and how the interconnectedness of the emotions affects their overall recovery process. Specifically, negative emotions such as anger, grief, and fear frequently dominate the initial recovery period, often impeding physical and emotional rehabilitation. Prolonged denial and grief can even lead to depression and identity loss, while fear, particularly of re-injury, also extends the recovery process. Conversely, optimism and acceptance, though less common, are emotions that foster motivation and cooperation with rehabilitation programs, ultimately leading to improved outcomes. Therefore, this study emphasizes the importance of additional research to develop early psychological interventions and supportive environments in mitigating negative emotional responses and promoting resilience.

1. Introduction

As the interest in sports grows among younger generations, more people aspire to become professional athletes, causing them to start playing at a young age (Yadava & Awasthi, 2016). While there is joy that comes with playing sports, there is just as much pain that comes with it. Over 90% of athletes get injured while playing sports—and most of the injuries are more severe than just a scratch or a sprain (Dacus et al., 2023). However, the moment of the injury is not the most painful moment during an athlete's career; the recovery process is both physically and mentally painful for many athletes because their job is deeply embedded in their identity, causing them to lose skill and confidence (Dacus et al., 2023). While there have been many studies conducted on the correlation between injuries and the mental health of athletes, most studies do not go in depth into the different psychological responses during the recovery phase (Smith, 2012). Therefore, this study aims to investigate how athletes may respond to different psychological reactions to an injury.

This literature review will focus on the most common psychological responses after a sports injury, specifically emotions such as anger or fear. The main research question being addressed in this study is: What are the common psychological responses athletes have after a sports injury, and how are they related to one another? The main purpose of this study is to identify the different emotional responses that different studies have identified and find a

correlation between the different emotions. While previous research highlights how each response is caused, there are gaps in the literature on how the responses are linked to each other, ultimately impacting the athlete to different degrees.

The review will be structured in a way that each emotion is introduced one at a time. Multiple literature about each psychological response will be analyzed and connected with each other to ensure a flow in the discussion. Then, another reaction that is related to the previous one will be analyzed in the same format. The four different categories of responses being analyzed are denial, grief, fear, and optimism. Following these sections, the review will identify the general amount of time taken to psychologically recover from an injury and solutions to how to recover. Finally, in order to connect the content of the review to a broader scope, implications such as how to improve past studies will be discussed in the conclusion.

This study is important in the field of psychology, as it gives opportunities to future researchers to study which emotions are the most common and why people have different responses. Therefore, this review is significant because it gives a comprehensive overview of the psychology behind sports injuries.

2. Background

Sports injuries are a major aspect of athletes' careers, as they often cause physical setbacks and profound psychological responses. This study serves an important purpose in the literature, as understanding these psychological responses and how they interrelate is essential for developing effective interventions to help injured athletes. However, when an athlete is injured the psychological responses to sports injuries are largely overlooked, with the focus primarily on physical rehabilitation because it is their job to continue playing (Smith, 2012). Through past actions to increase awareness of mental health, however, athletes and coaches have started to focus on the mental aspect of injury recovery as well. The most common responses to an injury, range from denial and anger to fear and acceptance, highlighting the complex and multifaceted nature of the psychological experience after an injury.

Psychological responses refer to the emotional, cognitive, and behavioral reactions an individual has due to external or internal stimuli; however, this review will focus mainly on the emotional reactions athletes have. One of the earliest and most famous frameworks in understanding these responses was the Kübler-Ross model of grief, which suggested that injured athletes often experience different stages of grief in a consistent pattern: denial, anger, bargaining, depression, and then acceptance (Avis et al., 2021). Although this model was initially developed for terminal illness, athletes also experience a similar set of emotions after an injury.

Although athletes are often noticed only for their performances in each game, they go through a complex relationship with psychological responses as well (Train, 2015). For instance, denial and anger can impede their progress toward acceptance, while fear can coexist with both grief and optimism, complicating the recovery process. Understanding how these responses interact is crucial for providing holistic support to injured athletes and making sure they return back mentally and physically stronger.

3. Methods

3.1 Search Strategy

This literature review was conducted to systematically analyze existing research on the psychological responses of a sports injury. The academic databases used to find relevant literature were PubMed and Google Scholar to ensure there was a wide variety of sources to read through. Studies were chosen from peer-reviewed and scholarly journals to ensure the research was reliable and informative. During the search, key words such as “psychological responses” and “athlete’s mental recovery” were used to find relevant research.

3.2 Criteria

Although there were no specific exclusion criteria for this study, sources were only included if they were in the English language, were published within the past 3 decades, and were written in reliable journals. In order to select relevant literature, the abstracts or summaries were read, followed by a full-text review.

3.3 Data collection

The specific information extracted from most literature included the author’s details, published year, study design, and key findings. The software, Zotero, was used throughout the process to organize and identify important data. Sources were found and examined in an order that corresponded to the structure and order of the paper. Once the data was collected, the literature was grouped according to the key findings and to match the flow of the literature review. However, limitations may exist in the review due to the search scope being limited to papers written in English.

4. Discussion

4.1 Anxiety and Anger

Injuries generally lead to a negative set of emotions, as they are a significant obstacle in an athlete’s career. When an athlete is injured, the most common response is stress and anger (Train, 2013). Train’s study collected data from 1000 athletic trainers through an online survey. The study’s results indicated that stress and anger are emotions that are very often encountered by athletes during the recovery process. Many of the athletic trainers in the study found that the emotions the athletes felt after an injury directly translated into their behavior by not adhering to the coaches or showing a negative attitude towards others. Another study by Yadava & Awasthi (2016) further highlights that although following a set rehabilitation process is important, it most often adds more stress to the athlete, causing them to feel more anger towards themselves and others. However, as the duration after an injury increases, athletes tend to transition from negative emotions to positive emotions (Brewer, 2017).

Nevertheless, there are still cases where athletes continue to stay in denial of their injury, leading to more serious issues such as high blood pressure or mental health issues, namely depression (Levine, 2009). Depression can lead to athletes losing motivation and stopping

physical activity (Marconin et al., 2023). This study investigated how Musculoskeletal injuries correlate with depressive symptoms through a systematic review. The study proved that negative emotions such as anxiety and anger can cause depression, and depression can cause tense muscles and distractibility, leading to higher risk of more injuries. Therefore, the most common emotions of anxiety and anger can lead to athletes becoming more distant from their sport.

4.2 Grief and hopelessness

Similar to anger, grief can also result in depression and other disorders. Grief has also been identified as one of the normal responses to a distressing situation such as an injury. While anger can lead to an athlete losing their motivation and joy to play a sport, grief can result in athletes losing their sense of identity (Evans & Hardy, 1995). Because athletes are so invested in the sport they play, it usually plays a role in shaping who they are as individuals. Similar to how people are in grief after they lose a loved one, losing the chance to continue playing a sport for a period of time can cause grief and sadness for an athlete. According to Evans & Hardy's research, grief, anxiety, and anger are closely intertwined as they appear very easily and can cause a detriment to one's physical health.

Furthermore, grief can be translated directly into a loss of self-esteem and a feeling of hopelessness (Park et al., 2023). Research has shown that when an athlete feels hopeless and down, it is very difficult to get them to cooperate with the rehabilitation process, causing a longer duration of recovery. The Kubler-Ross model further highlights that coping with an injury is a difficult process as it is difficult to accept the fact that they have lost a part of their athletic identity (Tatsumi & Takenouchi, 2014). In this study, the researchers examined emotional stability, temporal perspective, and team attachment through a survey that included a series of scales such as the Athletic Injury Psychological Acceptance Scale. The results showed that when an athlete was not emotionally stable, it affected both their temporal perspective and team attachment, causing them to feel more disconnected from their sport. Therefore, psychological interventions must be taken in order to prevent the loss and grief an athlete feels from distracting them from returning back to being an athlete.

Because athletes are admired and known by many people, they are often forced to keep their negative emotions hidden so that the public does not see the weak and hurt side of them. However, Tatsumi & Takenouchi's study concluded that if athletes constantly suppress their grief, it can lead to anxiety and depression as mentioned by Evans & Hardy's study. Therefore, previous literature has proven that it is common for athletes to experience grievance, but it can be harmful for them both mentally and physically.

4.3 Fear

Another response that leads to the recovery process being extended is fear. Throughout the entire recovery process, athletes have feelings of fear at multiple different stages. Initially, at the moment an athlete is hurt, they fear the intensity of their injury. Later on during rehabilitation, athletes fear that they may never recover properly. Lastly, once an athlete is

allowed to play again, they have a fear of re-injury (Hsu et al., 2017). The fear of re-injury and the fear of movement, kinesiophobia, are the most common among injured athletes, as they are now aware of the pain and struggles behind the rehabilitation process (Ambegaonkar et al., 2024). Ambegaonkar et al.'s study utilized a systematic review, identifying that athletes experience kinesiophobia both physically and mentally and that it can elicit feelings of anxiety. Another study by Hsu et al. utilized a clinical review to identify the impact of the fear of re-injury on an athlete's rehabilitation process. The study proved that when athletes experience more fear, their rehabilitation takes longer and is not as effective, preventing them from returning back to their sport in good conditions. Therefore, the fear that athletes feel can prevent them from being mentally stable, leading them to serious issues such as depression or not being able to safely return to their careers.

While fear is caused directly because of an injury, the levels of fear can differ due to other factors such as stress—high levels of stress correlate positively with risk of injury, leading to an athlete becoming more cautious (Pal & Kalra, 2021). Additionally, athletes feel fear from different objects and scenarios as well; in a questionnaire conducted by Ohji et al. (2021), athletes said they felt high levels of fear when they were in contact with another player, in contact with the ball, or were landing from a jump. Therefore, fear is the most common emotional response post-injury when an athlete is put back into a game.

4.4 Optimism and acceptance

Unlike the previous psychological responses mentioned, athletes sometimes feel positive emotions after an injury as well. However, these responses are not very common, and it takes a while for them to appear. Once athletes spend some time in rehabilitation, research found that they started to show decreased levels of depression and increased levels of acceptance and motivation after each session (Marconcin, 2023). Although accepting the fact that they are injured is very difficult, it is also the best solution to recovering faster. Williams et al.'s study found that when an athlete was optimistic, they were able to translate it into their rehabilitation process, being able to recover mentally and physically (2020). Thus, showing a positive emotional response after an injury is the most beneficial solution to an athlete.

4.5 Solutions

While an athlete may think it is impossible to recover from a severe injury, if they are mentally trained to show more positive emotional responses, they will be able to start playing again. As mentioned above, being accepting of one's injury and efficiently transitioning into gaining emotional stability will allow athletes to cooperate with the rehabilitation better (Tatsumi, 2023). Furthermore, the environment and people a recovering athlete is surrounded by play an important role in the speed of their recovery. Having supportive peers and family members as well as people with positive mindsets can motivate an athlete to try harder to recover (Park et al., 2023). While recovering is mainly up to the athlete's motivation, outside factors can influence how the athlete behaves and tries as well.

5. Conclusion

While sports are generally a source of joy and energy for athletes, it can also be a source of pain and stress. In order to reduce the amount of negativity caused by sports, it is important to research how to help improve the mental health of athletes, especially after an injury. As there is a limited amount of literature related to the connections between the different emotional responses, additional studies can be conducted in the future to identify the connection. Specifically, the studies can focus on the process of transitions between emotions during a recovery, acknowledging the fact that emotions are not static or permanent. Additionally, studies can be conducted to find the best solution to transitioning into feeling positive emotions or recovering from an injury in the shortest amount of time possible. Outside of the topic of emotional responses, additional research can be done on the impact of personal experiences and external factors on the recovery process.

Furthermore, the conclusions made in this study can be further applied to the field of cognitive psychology and topics such as memory. After an experience of being injured is incorporated in an athlete's memory, it can change their behavior and attitude, even outside of sports. Therefore, through the analysis of psychological responses, more research can be done in the future related to the cognitive behavior of people in general.

References

- Ambegaonkar, J. P., Jordan, M., Wiese, K. R., & Caswell, S. V. (2024). Kinesiophobia in injured athletes: A systematic review. *Journal of Functional Morphology and Kinesiology*, 9(2), 78. <https://doi.org/10.3390/jfmk9020078>
- Appaneal, R. N., Levine, B. R., Perna, F. M., & Roh, J. L. (2009). Measuring postinjury depression among male and female competitive athletes. *Journal of Sport and Exercise Psychology*, 31(1), 60–76. <https://doi.org/10.1123/jsep.31.1.60>
- Avis, K. A., Stroebe, M., & Schut, H. (2021). Stages of grief portrayed on the internet: A systematic analysis and critical appraisal. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.772696>
- Brewer, B. W. (2017). Psychological responses to sport injury. *Oxford Research Encyclopedia of Psychology*. <https://doi.org/10.1093/acrefore/9780190236557.013.172>
- Clement, D., Arvinen-Barrow, M., & Fetty, T. (2015). Psychosocial responses during different phases of sport-injury rehabilitation: A qualitative study. *Journal of Athletic Training*, 50(1), 95–104. <https://doi.org/10.4085/1062-6050-49.3.52>
- Clement, D., Granquist, M. D., & Arvinen-Barrow, M. M. (2013). Psychosocial aspects of athletic injuries as perceived by athletic trainers. *Journal of Athletic Training*, 48(4), 512–521. <https://doi.org/10.4085/1062-6050-48.3.21>
- Evans, L., & Hardy, L. (1995). Sport injury and grief responses: A Review. *Journal of Sport and Exercise Psychology*, 17(3), 227–245. <https://doi.org/10.1123/jsep.17.3.227>
- Hsu, C.-J., Meierbachtol, A., George, S. Z., & Chmielewski, T. L. (2016). Fear of reinjury in athletes. *Sports Health: A Multidisciplinary Approach*, 9(2), 162–167. <https://doi.org/10.1177/1941738116666813>
- Laurel, D., Christopher, C., Ciara, C., Gil, G., & William M, W. (2023). Impact of traumatic sports injury on an athlete's psychological wellbeing, adherence to sport and athletic identity. *Journal of Sports Medicine and Therapy*, 8(3), 036–046. <https://doi.org/10.29328/journal.jsmt.1001070>
- Marconcin, P., Silva, A. L., Flôres, F., Nunes, A., Lourenço, J. F., Peralta, M., & Minghelli, B. (2023). Association between musculoskeletal injuries and depressive symptoms among athletes: A systematic review. *International Journal of Environmental Research and Public Health*, 20(12), 6130. <https://doi.org/10.3390/ijerph20126130>
- Ohji, S., Aizawa, J., Hirohata, K., Ohmi, T., Mitomo, S., Koga, H., & Yagishita, K. (2021). Injury-related fear in athletes returning to sports after Anterior Cruciate Ligament Reconstruction - a quantitative content analysis of an open-ended questionnaire. *Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology*, 25, 1–7. <https://doi.org/10.1016/j.asmart.2021.03.001>
- Pal, S., kalra, S., & Awasthi, S. (2021). Influence of stress and anxiety on sports injuries in athletes. *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*. <https://doi.org/10.7860/jcdr/2021/45973.14702>

- Park, A. L., Furie, K., & Wong, S. E. (2023). Stronger athlete identity is a risk factor for more severe depressive symptoms after musculoskeletal injury in pediatric athletes: A systematic review. *Current Reviews in Musculoskeletal Medicine*, 16(5), 220–228. <https://doi.org/10.1007/s12178-023-09828-0>
- Smith, A. M. (2012). Psychological impact of injuries in athletes. *Sports Medicine*, 22(6), 391–405. <https://doi.org/10.2165/00007256-199622060-00006>
- Tatsumi, T. (2023). Individual-differences affecting emotion regulation behaviors of injured athletes: A retrospective quantitative study. *PLOS ONE*, 18(11). <https://doi.org/10.1371/journal.pone.0294551>
- Tatsumi, T., & Takenouchi, T. (2014). Causal relationships between the psychological acceptance process of athletic injury and athletic rehabilitation behavior. *Journal of Physical Therapy Science*, 26(8), 1247–1257. <https://doi.org/10.1589/jpts.26.1247>
- Williams, T., Evans, L., Robertson, A., Hardy, L., Roy, S., Lewis, D., & Glendinning, F. (2020). The role of optimism and psychosocial factors in athletes' recovery from ACL injury: A longitudinal study. *Frontiers in Sports and Active Living*, 2. <https://doi.org/10.3389/fspor.2020.00116>
- Yadava, V., Awasthi, P. (2016). Role of Psychological Factors in Management of Sports Injuries: An Overview. *Journal of Management Ethics and Spirituality*, 9.

The Effects of Message Framing on Charitable Donations From High School Students

By Ketan Shah

Abstract

Young adults are among the most committed to environmental issues. However, their financial support of organizations that are committed to environmental change is minimal (Harden, et al., 2020). The purpose of this study was to evaluate two types of messaging to assess their effectiveness in generating positive financial contributions. This study's survey was structured as a pre-and post-exposure design, where respondents were asked questions regarding their initial donation intent before seeing any message and, each time after viewing a message, were asked the same questions to see any change in desire to donate. The results of this study demonstrated that informational message frames were more effective in acquiring donations from high school students in comparison to messaging strategies that emphasize peer participation. These results can be used by charities to understand what message frames work the best, as well as ones that are not effective, so these organizations can support their cause from donations.

Keywords charitable donations, environment, message framing, donation intent, social proof, high school students, appeal, inclination

1. Introduction

With the rise of communications in the 21st century, charities are finding it increasingly more important to discover ways to make their communications more appealing and effective (Burton, 2022). It is likely that with more effective messages, not-for-profit organizations and institutions may increase revenue, donations. This study is focused on charities, and message frames can be more effective in motivating donation intent from the young market segment. The study is impactful because the demographic of respondents is Gen Z high schoolers who represent the future of environmental donations, so understanding what motivates them will affect the success of those organizations. It is also important because there have been no studies on the impact of social proof message frames, so this study will provide new and important information to charities that want to maximize donations. My research question is: Will social proof messaging strategies work best to engage a GenZ high school audience in intent to donate to an environmental cause? After conducting this experiment, it was discovered that informational message frames are more effective in eliciting donation intent and volume from high school students.

2. Literature Review

2.1. Improving the Effectiveness of Fundraising Messages: The Impact of Charity Goal Attainment, Message Framing, and Evidence on Persuasion

Previous research on the effectiveness of fundraising messaging (Das, et al., 2008) has indicated...

Three factors that could impact donation intent and volume. This research paired impact on donation intent with the type of expression of evidence: anecdotal vs statistical evidence. It was discovered that statistical evidence is more persuasive when paired with negatively framed messages. Anecdotal evidence is more effective when paired with positively framed messages. Furthermore, regardless of the message frame, subjects regarding donations for charity or “good” causes developed more donation intent. While this article was helpful in answering its question, it did not address the effect on donation intent from informative or social proof message frames. It also does not specifically investigate Gen Z, which provides a crucial segway to my research, as they are the financial future of donations.

2.2. The relation of approach/avoidance motivation and message framing to the effectiveness of charitable appeals (Jeong, 2010)

This study explored how individual differences in motivation, specifically through the Behavioral Activation System (BAS) and Behavioral Inhibition System (BIS), affect how people respond to persuasive messages. Individuals high in BAS, who are motivated by rewards and positive outcomes, were more receptive to gain-framed messages that highlight the benefits of taking action. In contrast, those high in BIS, who are sensitive to threats and punishment, responded better to loss-framed messages that emphasize the consequences of not acting. The researchers found that when the framing of a message matched a person’s motivational orientation, the message was significantly more persuasive. This supports the idea that tailoring messages to a person’s psychological disposition can increase their effectiveness. Despite this, this study did not look into the effects of informational or social proof message frames or the effects on Gen Z high schoolers, so my study will provide new insight to the research world.

2.3. Effects of age, need for cognition, and affective intensity on advertising effectiveness (Huhmann et al., 2009)

This study looked at how individual differences, such as age, need for cognition (NFC), and affective intensity (AI), influence responses to advertisements with various message frames - rational, positive-emotional, and negative-emotional. It was discovered that younger adults preferred the emotional appeals, whereas older adults showed more favorable responses towards rational messages. People high in NFC preferred the more emotional ads, and people low in NFC preferred more negative-emotional framing. Lastly, people with high AI showed more positive responses across all message framing. All of these findings show that customizing an advertisement to favor one’s traits and personality can improve message effectiveness. Once again, this study does not cover informational or social proof message frames or specifically address high schoolers, so my study will provide new and important information and is not redundant.

3. Question and Hypothesis

3.1. Research Question

It is essential that we discover the most effective ways to elicit donations. The research question addressed is: Will social proof messaging strategies work best to engage a high school audience in donation efforts?

3.2. Hypothesis

Based on previous research, the hypothesis for this study is that messages that emphasize social proof and peer participation will work best to engage a Gen Z high school audience in donation intent? Marketing messaging that is focused on other participating donors will be more effective in generating donations for a cause. The primary independent variable was the message framing, as the experiment started with no exposure and then used two different types of exposure. Other independent variables were centered around demographic information collected, such as gender and total annual household income. The dependent variable was consistent, however; this was donation intention.

4. Methodology

4.1. Sample: GenZ men and women

The survey participants were 63 high school students in Illinois, Michigan, California, and Massachusetts. This was a convenience sample from a suburban high school in a suburb of a large Midwestern city.

4.2. Experiment Method

This experiment tested whether or not message framing related to social proof increases donation intent amongst high school students. The data collection tool was a Google Forms survey, which was sent out to about 90 people, and 63 responses were collected, making the response rate approximately 7 responses per 10 requests. In this survey, the donation intent of the participants were assessed in two ways. One was a rating on a scale from 0 to 5, 5 being the strongest desire, on donation intent. The other assessment was asking the respondent how much money they would be willing to donate. To start, participants were asked questions to assess a baseline in terms of their knowledge on this topic and preliminary intentions. Next, participants were presented with an initial exposure message (view Appendix A) that contained an informative message frame. The participants were subsequently exposed to a second flyer (view Appendix B) that emphasized social proof and peer participation. This flyer attempted to persuade the audience by showing that many others were also donating and supporting this cause. The same two questions as from the first flyer were asked.

Demographic information on gender and total household annual income was also collected. This information was vital to the survey because it is to understand how donation

intent varied based on gender, and especially household income. The specific questions and flyers are included in the attached Google Forms survey.

The pre- and post-exposure design allowed for direct comparison of each subject's donation intention difference based on the contrast between having no exposure and two different types of exposure. Statistical analysis of the data was used to determine the change in donation intent before and after the exposure to the two differently framed messages. This approach will directly test the hypothesis that peer involvement messages in donation generate greater amounts of money.

5. Results

5.1. Results From All Respondents

View Appendix C. This graph represents all 63 respondents. After collecting the data and graphing, both sets of bars showed a peak in donation intention after seeing the first message and then a more significant decrease in intention after viewing the second message. In other words, the high schoolers were more motivated to donate to the environmental cause after seeing the first message, and after viewing the second, their intention to donate decreased. This means that the first message, the informational one, was more effective in eliciting donations and desire from high school students.

5.2. Results From Only Females

View Appendix D. This graph represents all of the female respondents. I chose to graph this information separately because I wanted to see if there was any difference in donation intent or reaction to message framing across genders. The data was graphed the same way as the previous one. Once again, both sets of bars showed a peak in donation intention after seeing the first message and then a more significant decrease in intention after viewing the second message, meaning that the informational message was more useful than the social proof one.

5.3. Results From Only Males

View Appendix E. This graph represents all of the male respondents. I chose to graph this information separately because I wanted to see if there was any difference in donation intent or reaction to message framing across genders. The data was graphed the same way as the previous two. Even though both sets of bars showed a peak in donation intention after seeing the first message and then a more significant decrease in intention after viewing the second message, the male respondents had much bigger differences in intent and donation amount across the three periods of questioning in comparison to the females. In other words, the results from the male participants were more varied than those of the females.

5.4. Results From Respondents From Households With Annual Income Over \$150K

View Appendix F. This graph represents all of the respondents who come from a household that makes a total annual income of \$150,000 or more. I chose to graph this information separately because I wanted to see if the amount of money one makes would make an impact, specifically on the donation amount question. The data was graphed the same way as the previous one. Once again, both sets of bars showed a peak in donation intention after seeing the first message and then a more significant decrease in intention after viewing the second message. However, this graph showed the most consistency in terms of the amount of people willing to donate more than \$11 to the cause after all three questioning periods. This most likely hints to the impact that the household income has on the responses.

6. Conclusions and Discussion

6.1. Conclusions

Analysis of data in this experiment contradicts the hypothesis, which was that message framing focused on social proof will elicit more donation intent from Gen Z towards a charitable cause. This experiment indicated that the informational message frame, not the social proof exposure, was the most convincing, as the majority of people felt the most desire to donate after viewing the former. There was a spike in donation intent and desired donation amount after viewing the first flyer, which was informationally framed. Then, after the respondents viewed the second flyer, which was framed using social proof tactics, there was a significant decrease in donation inclination. This conclusion was consistent across all demographics, as the gender or level of household income did not change the overall result.

6.2. Discussion and Directions For Future Research

One strength included having a convenience sample - I was able to use my high school class as survey participants, which made data collected practical and simple. Additionally, because I only included two stages of exposure and only assessed the effects of two different types of message framing, the survey was not too lengthy, which allowed respondent attention levels to remain high throughout the questioning. This ultimately led to more truthful answers, as a survey too long could lead to respondents mindlessly clicking answers without understanding the impact of their answers. However, a limitation faced during this study was that I focused this research solely on high school students. This shrunk the amount of potential respondents by a lot and less respondents leads to more margin for error. In other words, I had to accept that there would be some margin of deviation from the real truth due to having a smaller sample size. Another limitation faced was that I did not assess all possible message frames, so my data can only show that an informational strategy is more effective in comparison to a social proof strategy. I can't conclude that an informational strategy is the most effective, as there are many other methods or combinations of message frames that could be more useful.

Despite this, it is clear that making messages that emphasize peer participation and social proof may not be the best method in gaining donations from high school students, as making

flyers that are more informational appears to be more effective. Therefore, organizations trying to obtain as many donations as possible from a younger generation should consider using informational message frames, as this is more productive. As mentioned earlier, an informational messaging strategy may be more effective, but it might not be the most useful. There are many other factors of a flyer or message that can be changed that could alter the results, which leaves plenty of room for further research that can have a positive effect on donations that can support important environmental causes.

Works Cited

- Burton, Aliyah. "The Intersection of Digital Communication & Non-Profit Organizations." Liberty University, vol. ., no. ., Oct. 2022.
- Das, Enny, et al. "Improving the Effectiveness of Fundraising Messages: The Impact of Charity Goal Attainment, Message Framing, and Evidence on Persuasion." *Journal of Applied Communication Research*, vol. 36, no. 2, May 2008, pp. 1–15. Research Gate.
- Harden, Josh, et al. "An Insight Into the Charitable Giving of Young Adults & Students." *Reason Digital*, vol. ., no. ., Sept. 2020.
- Huhmann, Bruce, et al. "Effects of Age, Need for Cognition, and Affective Intensity on Advertising Effectiveness." *Journal of Business Research*, vol. 64, no. 1, Sept. 2009, pp. 1–6. Academia.
- Jeong, Esther, et al. "The Relation of Approach/Avoidance Motivation and Message Framing to the Effectiveness of Charitable Appeals." *Social Influence*, vol. 6, no. 1, Nov. 2010, pp. 1–8.
- Shah, Ketan. "Environmental Messaging Survey." Google Forms, 7 Apr. 2025, <https://forms.gle/NYkHviYL8YGTyv336>.

CRISPR Cas-9 as a Potential Cure for Wilson's Disease By Satoshi Toya

Abstract

Wilson's Disease is an inherited copper metabolism disorder that is characterized by the pathological accumulation of copper throughout the body. It is caused by a mutation in the ATP7B gene – a gene that encodes a transmembrane copper-transporting protein – which inhibits the body's ability to transport copper properly. This results in the accumulation of copper in organs such as the brain and liver, which cause ophthalmological, hepatic, neurological, and psychiatric symptoms. The primary methods of treatment are chelators and zinc salts, which lower copper levels through different mechanisms; these treatments can become expensive because they need to be used continuously throughout the patient's lifetime.

However, CRISPR – an immune system used by various bacteria and archaea to eliminate invading viruses and plasmids – can be utilized as an effective gene editor, with many possible applications, including the treatment of chronic genetic diseases. The most prominent version of CRISPR used for this purpose is the CRISPR-Cas9 system. This paper aims to achieve multiple goals: to explain Wilson's Disease, mechanisms, and symptoms; to overview CRISPR and how CRISPR-Cas9 can modify DNA in genes; and to propose how CRISPR could be used to potentially permanently cure chronic genetic diseases, such as Wilson's Disease.

Introduction

Wilson's Disease is an inherited disorder of copper metabolism that causes abnormal copper accumulation in many organs, particularly the liver and brain, resulting in a wide range of symptoms^{5, 121}. This copper overload can result in a myriad of negative consequences, including oxidative stress and protein toxicity, most notably in the mitochondria.

On a molecular level, Wilson's Disease arises due to oxidative stress, copper-mediated toxicity occurs when excess dissolved copper catalyzes the formation of hydroxyl radicals in the presence of peroxides via Fenton reactions¹⁴⁸.

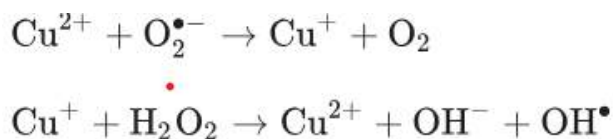


Fig 1: Fenton reaction

Hydroxyl Radicals are strong oxidants that can impede protein functionality by reacting with amino acids such as tyrosine and cysteine². This substance not only directly interacts with DNA and RNA and causes oxidative DNA damage⁵⁸, but it can also interact with fatty acids, which in turn can create lipid peroxides and aldehydes that can react with DNA or amino acids to form adducts^{48, 57, 82}. The adducts themselves are also detrimental to the patient's health as DNA adducts, for example, cause the DNA part to be damaged, which leads to its abnormal replication in the body and could cause mutations that potentially lead to conditions such as cancer¹²⁸.

Copper can directly target specific proteins in the body, most notably membrane proteins, by binding to thiol residues, resulting in the cross-linkage of these proteins¹⁷⁸. A possible result of this phenomenon is the mitochondrial membrane permeability transition(MPT), which is caused by copper directly interacting with thiol groups in important mitochondrial proteins such as the adenine nucleotide translocase^{42, 159}. Additionally, MPT can be present in organs such as the liver and brain, although the severity of MPT caused by copper varies from organ to organ¹².

Wilson's Disease Pathophysiology

Wilson's Disease is mainly caused by mutations affecting the ATP7B gene, which contains 20 introns and 21 exons, and is located on the long arm of chromosome 13 (Specifically region 13q14.3) with a genomic length of 80 kb^{119, 185}.

The protein is first synthesised in the endoplasmic reticulum, and then it is relocated to the Trans-Golgi-Network (TGN) within hepatocytes. The ATP7B gene itself is most highly expressed in the liver, but it is also found in other organs such as the kidneys, brain, and lungs¹⁸⁰.

The Genetics of Wilson's Disease

The ATP7B protein itself consists of 1465 amino acids (165 KDa) which are organized into many different parts of the protein: the phosphatase domain (A-Domain), which is identifiable by the amino acid motif TGEA alongside TGDN and GDGVND; the phosphorylation domain (P-domain), identified by the amino acid motif DKTGT; the M domain, which consists of 8 transmembrane ion channels; the nucleotide binding domain (N-Domain), identified by the amino acid motif SEHPL; and the N-Terminal metal binding domain (MBD), which contains 6 copper binding sites with the amino acid motif GMXCXXC and plays a central role in accepting copper from the copper chaperone ATOX1^{20, 33, 87, 138}.

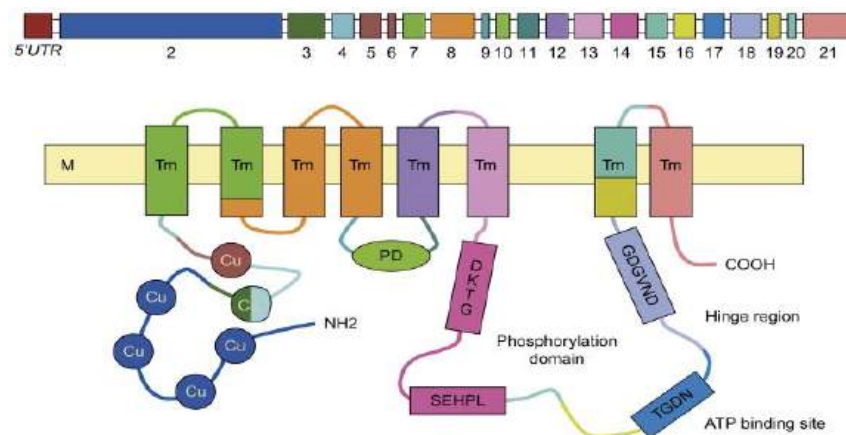


Fig 2: The general structure of the ATP7B protein
(The genetics of Wilson disease - PMC)

ATP7B is responsible for encoding a copper cation-transporting P-type ATPase that facilitates the transportation of copper between cells. These proteins are upregulated in liver

hepatocytes, and the proteins are mainly found in the Golgi Apparatus^{36, 157}. This membrane-bound transporting protein facilitates the transport of copper from the cytoplasm to the Trans-Golgi Network, where copper binds to the protein apoceruloplasmin, forming ceruloplasmin, a copper transport protein. Ceruloplasmin then enters the bloodstream to distribute copper to the rest of the organs for their necessary function^{36, 91, 98, 139}.

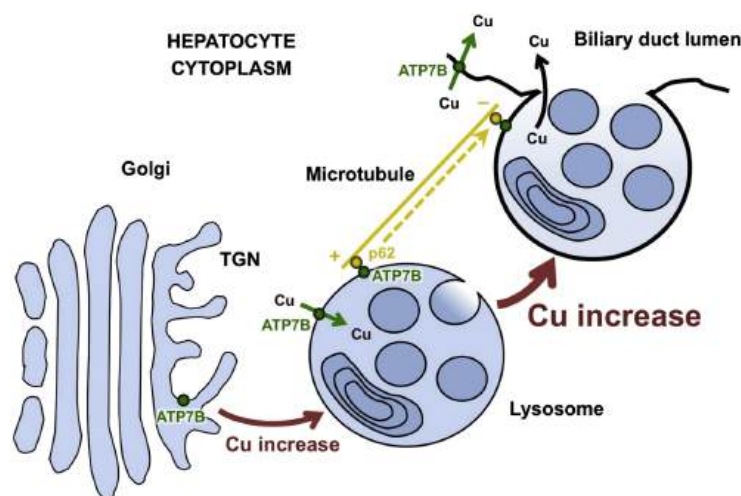


Fig 3: The general arrangement of ATP7B within the hepatocytes and its role in transporting copper

Focusing on the protein mechanism in greater detail, the active transport of copper across cell membranes begins with ATP7B binding copper at the MBD domain and transporting it across cellular membranes, driven by the hydrolysis of adenosine triphosphate (ATP). Then, the free copper binds intracellularly to GG motifs in the MBDs, followed by transport to the Cys-Pro-Cys (CPC) sequence motifs in MBD 6. Finally, the dephosphorylation of acyl-phosphate at the A-domain discharges copper across the cellular membrane¹⁸⁰.

Copper is necessary for bodily function in many ways, notably acting as a cofactor for many critical cuproenzymes essential for human metabolism¹⁵⁶.

An example of this is in Cytochrome-c oxidase, located in the inner mitochondrial membrane, which has two copper sites in the form of CuA and CuB³⁷. Cytochrome-c oxidase acts as the terminal oxidase of the mitochondrial respiratory chain in most aerobic organisms, transporting electrons from donors to a final acceptor to create ATP, simultaneously reducing molecular oxygen (O₂) into water (H₂O)³⁷.

Another example is human SOD1, a zinc-copper metalloenzyme in all cells that regulates the body's antioxidant response^{41, 107}. Functionally, SOD1 catalyzes the disproportionation of superoxide anions (O₂⁻) into molecular oxygen and hydrogen peroxide, which are further eliminated by catalase, glutathione peroxidases, and peroxiredoxins⁴¹. This biochemical process is important because without this type of regulation, the superoxide can react to produce more dangerous Hydroxyl radicals (•OH) that can cause further cellular damage^{100, 107}. Additionally,

copper is necessary for SOD1 to catalyze the disproportionation, as reduction of superoxide occurs at the copper-zinc metal center^{4, 26, 171}.

Another example of a necessary copper-containing protein is ceruloplasmin, which serves as both a copper transport protein and a ferroxidase with the ability to oxidize Fe(II) to Fe(III). Oxidation of the iron substrate allows iron to bind more strongly to transferrin, which is responsible for iron transport and distribution in the body^{38, 118, 156}. Ceruloplasmin also removes excess copper in hepatocytes through vesicles, where ceruloplasmin is excreted through the apical membrane to create bile^{98, 139}.

However, when the *ATP7B* gene is mutated, it causes a mutation in the copper-transporting P-type ATPase that leaves the enzyme unable to function properly¹⁰⁸. As a result, this condition leads to both the impaired incorporation of copper into ceruloplasmin³³ and the excretion of copper into bile. Copper accumulates in the liver over time, in some cases increasing copper concentrations from 5 - 20 times normal copper levels in the body²⁷. When the liver reaches its copper-carrying capacity, excess copper is released into the bloodstream and travels back to other organs, resulting in an array of problems¹⁶¹.

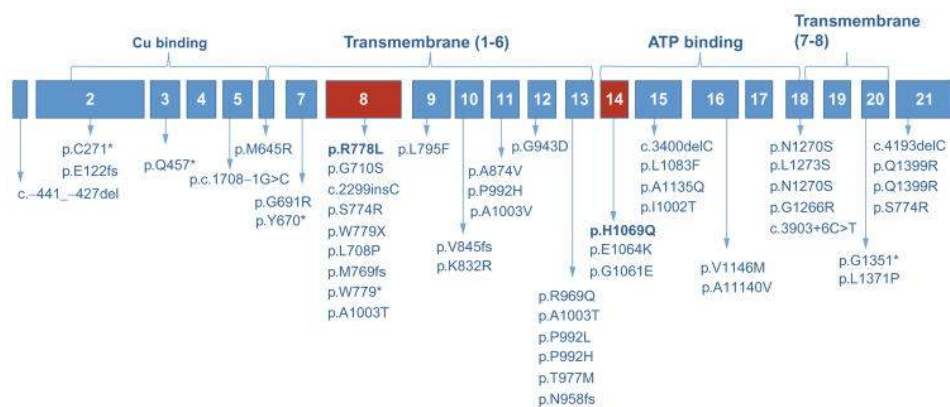


Fig 4: A schematic of the *ATP7B* gene with common mutation sites
(The genetics of Wilson disease - PMC)

One of the most common mutations that leads to the disease is the missense mutation H1069Q in exon 14. It's common in Central, Eastern, and Northern Europe, where approximately 50–80% of Wilson's Disease patients carry at least one H1069Q allele inherited from their parents³⁶. In this mutation, the histidine of the conserved SEHPL motif in the N-terminal domain of the *ATP7B* protein is replaced with glutamic acid, which results in N-domain protein misfolding, abnormal phosphorylation in the P-domain, and decreased ATP binding affinity¹³². This mutation can also lead to decreased thermostability and the abnormal localization of the protein to the Trans-Golgi-Network, further disrupting transport processes¹²⁹.

Symptoms

There are multiple symptoms related to Wilson's Disease.

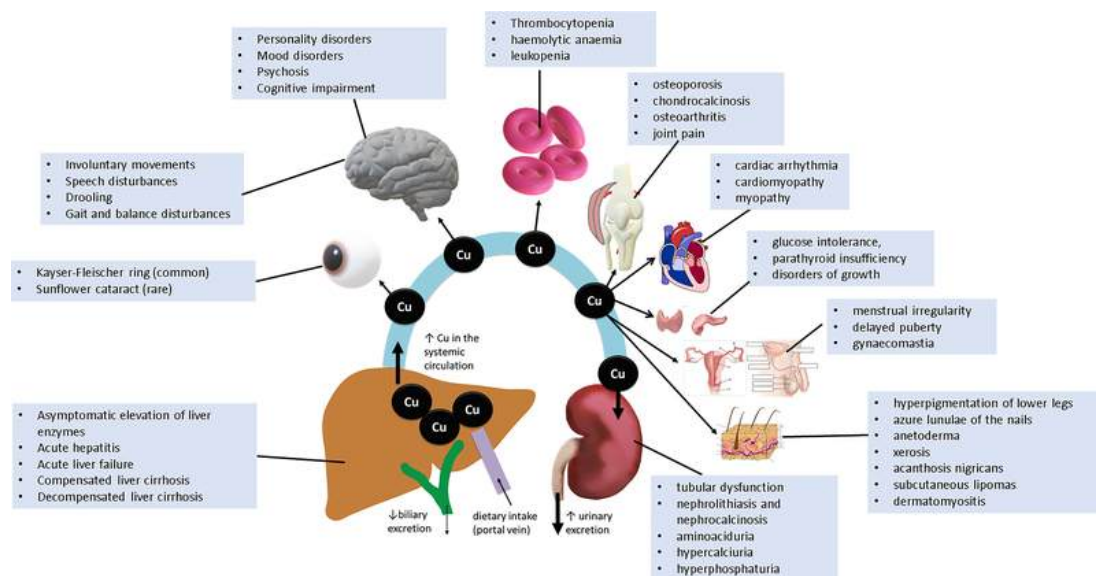


Fig 5: A listing of the organs generally affected by the excess copper caused by Wilson's disease. (Nature Reviews Disease Primers article: Wilson disease - PMC)

Hepatic Symptoms

Early symptoms of Wilson's Disease usually manifest in the liver, due to excess copper buildup in hepatocytes^{5, 168}. The severity of excess copper's effect on the liver can vary due to the lack of an exact genotype-phenotype correlation to Wilson's Disease. This can result in relatively small manifestations, such as a self-limiting Hepatitis-like inflammation, and in more severe cases, cirrhosis, or the scarring of the liver, and even acute liver failure¹⁶⁸. Furthermore, cirrhosis leads to jaundice, which can serve as a visual symptom in patients with liver damage. Without treatment, chronic liver disease worsens alongside the appearance of conditions such as portal hypertension, hepatosplenomegaly, ascites, low serum albumin concentration and coagulopathy²⁷.

The severity of these hepatic diseases that can cause cirrhosis is usually reflected by the Model for End-Stage Liver Disease(MELD)⁶⁹alongside Child-Pugh scores¹²⁷. And normally, these are determined by using abdominal ultrasound (US), CT, and MRI at diagnosis. The most common findings through these methods are events such as fatty infiltration, contour irregularity, and atrophy of the right lobe of the liver²⁷. For patients with chronic liver disease, especially those with cirrhosis, screening examinations for signs of portal hypertension and hepatobiliary malignancies¹²⁰ should be performed regularly¹⁰ and repeated depending on clinical status.

Neurological Symptoms

Neurological symptoms are the second most frequent behind hepatic symptoms and, on average, will onset around the age range of 20-30 (though outliers do exist)^{32, 121}. The most predominant manifestations are tremor, dystonia, and Parkinsonism—all of which are normally associated with dysarthria, gait and posture disturbances, drooling, and dysphagia^{93, 96, 168}. One of the most characteristic neurological symptoms is tremors (shaking or trembling movements),

which is a symptom described by about 55% of patients at diagnosis. Tremors can be either resting, postural, or kinetic, and movement is either unilateral or bilateral and occurs primarily in the arms. Over time, tremors can spread to the legs, head, and throughout the entire body²⁷.

Dystonia, or involuntary muscle contracting, is reported as one of the first Wilson's Disease symptoms in 11–65% of patients²⁷. It can involve a single body part, one segment, multiple segments like a face and a leg, or even be generalized throughout the entire body. The most characteristic representation of dystonia is an abnormal facial expression, or risus sardonicus, which consists of a fixed smile due to dystonia of the risorius muscle.

Ophthalmological Symptoms

Ophthalmological symptoms, predominantly Kayser-Fleischer Rings and sunflower cataracts^{28, 169}, both of which are caused by pathological copper accumulation in the eyes.



Fig 6: Kayser-Fleischer Ring viewed on the outside)
(Kayser-Fleischer rings | Radiology Reference Article | Radiopaedia.org)

Kayser-Fleischer Rings (KFRs) are ring-shaped copper deposits within the eyes that develop when copper is deposited throughout the cornea; a sulfur-copper complex—found only in the Descemet's membrane—then produces visible ring-shaped copper deposits in that area¹²¹. Specifically, the deposit is formed in the anterior chamber angle within the internal corneal layer of the Descemet's membrane, at Schwalbe's line^{15, 74}.

Normally, Kayser-Fleischer rings are present in almost all people with Wilson's Disease who have developed neurological or psychiatric symptoms. However, there are exceptions, with some patients not having KFRs despite experiencing neurological symptoms¹³⁴.

And generally, the presence of these copper deposits can alter the corneal subbasal nerve plexus(SBNP)—a monolayer of fine, amyelinated nerve fibers in the cornea. One clinical study demonstrated that Wilson's Disease patients can experience corneal neuropathy as a result of this monolayer accumulation. This was determined by an evaluation of Wilson's Disease patients' eyes using corneal confocal microscopy: factors like the number of nerve fibers, nerve fiber length density, number of nerve branchings, and number of beadings were all significantly lower when compared to those unaffected by Wilson's Disease. Additionally, in Wilson's Disease patients, the fiber tortuosity was higher compared to unaffected people, indicating that there was

a lot of nerve regeneration, an indication that nerve fibers in the eyes had died off as a result of neuropathy^{68, 149}.

Kayser-Flescher Rings become visible in the superior aspect of the cornea, followed by the inferior aspect, with subsequent filling in of the medial and lateral aspects. As a result, one would have to expose the entire cornea by lifting the eyelid to properly look for KFRs¹⁵⁰.

Furthermore, KFRs are detected and diagnosed using methods such as slit lamp examination(SLE), which makes KFRs appear as a golden-brown, golden-green, green-yellow, golden-yellow, bronze or reddish-brown coloring ring in the limbic area of the cornea³⁴.

Sunflower cataracts are characterized by copper deposition in the lens capsule, distinct from the lens cortex or nucleus, forming a central disk with radiating petal-like spokes that give it its name¹²⁵. Additionally, Sunflower cataracts seem to be the result of the accumulation of various compounds in the third posterior of the lens' anterior capsule, which includes copper, sulfur, and binding-copper proteins¹⁸¹.

Sunflower cataracts typically do not impair vision and are usually detected through slit lamp examination, as they are not visible to the naked eye or an ophthalmoscope⁹⁶.

Psychiatric Symptoms

Psychiatric Symptoms appear frequently in Wilson's Disease patients³². A 2014 study suggested that psychiatric symptoms can occur before, in tandem with, or after the diagnosis and treatment of Wilson's Disease¹⁷⁷. It can appear in younger patients as declining school performance, impulsiveness, and inappropriate behavior^{142, 152}; It can often appear in older patients as paranoia, schizophrenia, or depression-but it can also commonly appear as behavioral changes³².

The most common psychiatric symptom of Wilson's Disease is mood disturbances, with about 20–60% of Wilson's Disease patients developing depression, with a high rate of suicidal attempts ranging between 4–16% of WD patients^{146, 151, 152, 177}. Adding on to this, behaviors such as emotional lability, irritability and aggression, shallow cheerfulness, euphoria, social disinhibition, hypersexuality, lack of criticism, and deficits in planning and anticipating social consequences can be due to the lesions of the frontal lobe or its pathways¹⁹.

Additionally, Behavioural and personality disorders are also a common psychiatric aspect of Wilson's Disease, with the most common manifestations being irritability, aggression, and antisocial behaviour¹²². Other psychiatric conditions, such as catatonia, anorexia nervosa, bulimia, obsessive-compulsive disorder, and attention-deficit hyperactivity disorder, have also been reported in Wilson's Disease patients^{53, 83}.

Additional Symptoms

There are still more different manifestations of Wilson's Disease. However, the effect of Wilson's Disease on other parts of the human body has been less investigated compared to organs such as the liver and brain²⁷.

One manifestation is the negative effects that the excess copper from Wilson's Disease can cause to the skeletal system, which includes osteomalacia(softening of bones) and osteoporosis(weakening of bones), with increased incidence of spontaneous fractures being observed¹⁶⁵. Additionally, the copper can negatively affect the joints, particularly larger ones, with its accumulation in the synovial membrane and cartilage being suggested as the major cause of osteoarthritis(erosion of cartilage) and accelerated degenerative changes with deformities¹⁰⁹.

Another type of manifestation is cardiovascular symptoms, with myocardial copper accumulation being able to cause cardiomyopathy(decrease of heart muscle function) and arrhythmias(irregular heartbeat)²⁷. Additionally, pathological cardiac examination done on Wilson's Disease patients has shown symptoms such as interstitial and replacement fibrosis, intramyocardial small vessel sclerosis, and focal inflammatory cell infiltration¹⁶.

Some manifestations can affect the red blood cells in Wilson's Disease patients with the excess copper in the bloodstream being able to cause symptoms such as Coombs-negative haemolytic anaemia; this alongside other symptoms may be caused through the reaction of copper with membrane lipids, or copper inhibiting sulphhydryl groups of glucose-6-phosphate dehydrogenase and glutathione reductase which all can reduce cellular antioxidant capacity which leads to oxidative damage of the haemoglobin and cell membrane in red blood cells^{40, 160}.

Last but not least, there are also renal manifestations that are caused by the renal tubular epithelium filtering the non-ceruloplasmin-bound copper in the blood that can cause excess copper buildup in the renal parenchyma, leading to renal tubular dysfunction¹⁷⁶.

Wilson's Disease Incidence / Prevalence

Initially, it was believed that the universal prevalence of Wilson's Disease ,except for a select few small isolated populations, was 1:30,000, based on a study done by Scheinberg and Sternlieb in 1984¹³⁶.

More recent data from the U.S, Europe, and Asia suggest a prevalence range of about 1:30,000-1:50,000, clarifying the 1984 study results¹³⁶. Prevalence is relatively greater in Pakistan and India, due to factors like genetic consanguinity.

Wilson's Disease Treatment Available

Once a patient has been definitively diagnosed with Wilson's Disease, treatment is provided for a patient's entire lifetime^{35, 131}.

Chelating Agents

One type of treatment for Wilson's Disease is pharmaceutical treatment that can induce cupriuresis, or the urinary excretion of copper. This is done primarily with copper chelating agents, such as D-penicillamine and Trientine, or using zinc salts that decrease the absorption of copper in the digestive tract.

D-Penicillamine functions as a chelator that binds divalent metal ions like copper with its sulfhydryl functional group (–SH) to form a water-soluble complex that is excreted out of the body through urine⁴⁷.

Trientine is another copper chelating agent and can act as an alternative to D-penicillamine, particularly for those who are unable to handle D-penicillamine. Trientine binds to copper in a 1:1 ratio to form a strong, stable complex that is excreted in the urine.

D-penicillamine and trientine result in similar patient responses; However, trientine was more likely to exacerbate the neurological effects of Wilson's Disease compared to D-penicillamine⁴⁷. Conversely, D-penicillamine treatment is more likely to be discontinued due to side effects when compared with trientine¹⁶⁶.

When considering overall performance, trientine and D-penicillamine are not inferior to one another, as evidenced by a recent randomized open-label non-inferiority trial¹⁴¹.

However, these types of treatments can be expensive. An example of this is the price for the 2 drugs around 2015, where trientine(Syprine) was \$200 per 250-mg capsule and D-penicillamine(Cuprimine) costing \$55-\$60 per 250-mg capsule¹⁸². Around 2022, the annual cost for treatment was about \$78,000 for trientine, and from \$50,000 to \$200,000 for penicillamin³.

Zinc Supplements

There are also Zinc supplements, which can have multiple versions such as zinc acetate, zinc gluconate, and zinc sulfate. However, there is no statistically significant difference between these variants in terms of the improvement of liver function¹⁸. Zinc preparations work by inducing the synthesis/creation of metallothioneins while promoting the binding of copper to these metallothioneins in the enterocyte of the small and large intestines. The copper-containing enterocyte is then disposed of through it shedding naturally into the gut lumen, then it accompanies the rest of the waste out of the body. This overall process hinders the body's absorption of copper, and prevents the further exacerbation of Wilson's Disease from increasing⁴⁷. Its effects are generally slow as a de-copperating agent, with its effects happening over time rather than manifesting as a sudden, noticeable event. As a result, it's not suitable for treating more florid symptoms such as cirrhosis¹⁴⁰. Wilson's Disease still shows progression in the first few months of treatment¹⁴. When compared to the other types of treatment for Wilson's Disease, this is the cheapest one out of all of them, with the annual cost from 2022, with \$100 worth for nutriment types, and from \$550 to \$3650 for pharmaceutical-grade Zinc³.

Liver Transplant

As a more complex and surgical option, liver transplantation also counts as a treatment for Wilson's Disease^{1, 30, 167}. This approach is based on the concept that Wilson's Disease results from gene defects, such as those in ATP7B, which impair the liver's ability to manage excess copper. Consequently, a liver transplant can be viewed as a phenotypical correction of the gene, thereby restoring copper homeostasis. Normally, liver transplantation in Wilson's Disease

patients is suggested in cases with acute liver failure or decompensated liver cirrhosis with a lack of pharmacological treatment effect. The need for a liver transplant in cases of acute liver failure should be based on a specific WD clinical scale, such as the revised King's prognostic Wilson Index³⁰. On another note, the idea of whether uncontrolled neurological disease constitutes a need for liver transplantation as the “last chance” treatment option is highly controversial^{1, 30, 167}. The problems with this treatment, however, are that it can give rise to a myriad of long-term medical complications in the patient as a result of the transplant. This can include various conditions: cardiovascular complications that account for 20% of the deaths of the patients surviving at least 3 years from a liver transplant¹²⁶; Kidney disease where up to 80% of patients 5 years after a liver transplant will develop chronic kidney disease⁵⁰; Hyperuricemia which has been reported in 14–47% of post Liver transplant patients¹¹⁶, and may be related to deteriorating renal function due to acid uric having a high renal excretion^{95, 113}; Bone complications where 65% of Liver transplant recipients were affected by Atraumatic bone fractures because of cholestatic disease⁹; and many other possible side effects.

Additionally, liver transplants themselves can cost a lot of money, with the operation itself costing about \$600,000, not including the mandatory preoperative evaluations, which can cost from \$168,000 to \$308,000, and the postoperative medications that cost from \$671,000 to \$949,000 in the 10 years that follow the procedure⁵⁵.

When considering overall treatment, the average healthcare cost for a Wilson's Disease patient, based on a 2021 study on the cost burden for Wilson's Disease patients, was about \$3,887 per month after a Wilson's Disease diagnosis¹³⁵. And when looking at a larger scale, the total cost for Wilson's Disease hospitalizations was \$20.90, \$27.23, \$24.20, and \$27.25 million US dollars for the years 2016, 2017, 2018, and 2019, respectively⁸⁶.

CRISPR History

CRISPR (or clustered regularly-interspaced short palindromic repeats) describes a family of DNA sequences found in bacteria and archaea that facilitate an adaptive immunity against mobile and exogenous genetic elements such as viruses. Since their characterization by Emmanuelle Charpentier and Jennifer Doudna, CRISPR systems are now utilized as a precise gene-editing tool that has revolutionized the field of molecular biology^{6, 23}.

CRISPRs were first detected in the *E. coli* bacteria's genome in 1987, where they were characterised as unusual insertion elements consisting of a series of 29 nucleotide repeats separated by 32 nucleotide “spacer” sequences, all of which had appeared whenever bacteria came into contact with invading phage DNA⁶⁰. At the time, scientists were unable to determine the function of these sequences due to the lack of sufficient DNA sequence data, especially for mobile genetic elements⁶¹. In 1993, CRISPRs were first observed in the archaea *Haloferax mediterranei*¹¹¹, and they were subsequently detected in an increasing number of bacterial and archaeal genomes as gene sequencing became cheaper and more widely available in the life sciences. All of these example sequences were conserved, which helped to further the understanding of their function and importance.

Research related to CRISPR came to a head in the early 2000s when two parallel breakthroughs greatly accelerated genetic research efforts.

One breakthrough was an increased understanding of the function of CRISPR as an immune defense after the discovery of the sequence similarities between the spacer regions of CRISPRs and the sequences of bacteriophages, archaeal viruses, and plasmids were reported on by three different research groups^{11, 112, 124}.

The other major discovery was that several genes previously thought to encode DNA repair proteins for hyperthermophilic archaea were actually strictly associated with CRISPR and were designated as Cas (CRISPR-associated) genes as a result^{62, 102}. Comparative genomic analysis had suggested that CRISPR and the Cas proteins both work together to constitute an immune system to protect prokaryotic cells against invading viruses and plasmids (DNA-based threats), comparable to the eukaryotic RNA interference (RNAi) system¹⁰³.

A 2007 experiment proved this hypothesis using the lactic acid bacteria *Streptococcus thermophilus*, where the molecular mechanism of the adaptive immune response to a phage infection was explained⁶.

Additionally, it was demonstrated that CRISPR sequences are transcribed into RNA that is then cleaved and loaded into CRISPR–Cas proteins, resulting in an RNA–protein complex sufficient for RNA-guided double-stranded DNA endonuclease activity^{44, 66}.

CRISPR Mechanism

Specifically, CRISPR consists of two major parts. First, a CRISPR system must contain a genome with an entire CRISPR locus that consists of an array of short direct repeats interspersed with spacer sequences where the spacers and repeats are both uniform in length, but the repeats are identical in sequence content, and the spacers are highly variable in sequence content⁷³. Second, it also consists of Cas enzyme-producing genes organized in operon(s) to allow transcription by the cell^{7, 158, 170}.

Regarding the CRISPR locus in different species of bacteria and archaea, the repeats are about 32 base pairs (bp) long on average, and the spacers are generally of a similar size in the range of 20-72bp^{49, 52, 110}. Related species may have similar repeat sequences, but there is great diversity in the sequences of both spacers and repeats in bacteria and archaea⁸⁴.

In the original immune defense system, the CRISPR–Cas mechanism of action consists of 3 primary stages: adaptation, expression, and interference. In the adaptation step, a complex of Cas proteins encounters a protospacer-adjacent motif (PAM) and causes 2 double-strand breaks in the PAM's invading DNA molecule. The resulting fragment of foreign DNA (termed protospacer, is the target) is then integrated into the CRISPR locus between 2 repeats of the CRISPR array and becomes a spacer around the proximal end of the CRISPR array³¹.

In the expression stage, the Cas gene's expression, which produces Cas proteins, and the transcription of the repeat-spacer element from the protospacer, which produces a precursor CRISPR RNA (pre-crRNA), in which proteins and accessory factors process the pre-crRNA into short mature crRNA, all occurs³¹.

Finally, in the interference step, the crRNA and Cas proteins work together to recognize and mediate the cleavage of the foreign DNA to protect the host cells from the infection^{6, 103}.

Classification of CRISPR/Cas9 Systems

CRISPR/Cas systems are divided into 2 main classes, 6 main types, and 33 subtypes. Class 1 contains multiprotein effector complexes, while class 2 is defined by a single, multidomain, multifunctional effector protein⁶⁷.

All types are distinguished by the different structures of the effector modules, which contain unique signature Cas proteins. Additionally, each type is also classified into numerous subtypes that are characterized by the subtle differences in CRISPR locus organization and encode subtype-specific Cas proteins^{56, 81, 145}.

Overall, the primary features that define the type and subtype of the CRISPR/Cas systems are the Cas genes and the proteins they encode, which functionally and genetically differ across the different variants, as shown in the various steps of CRISPR-mediated immunity.

For example, the type I and III systems(both in Class 1) both produce specialized Cas endonucleases, which process the pre-crRNAs, which mature into crRNA and assemble into a large multi-Cas protein complex capable of recognizing and cleaving nucleic acids complementary to the crRNA³¹.

Type II(which is Class 2) systems, by contrast, process pre-crRNAs with a different mechanism where a trans-activating crRNA (tracrRNA) complementary to the repeat sequences in pre-crRNA triggers pre-crRNA maturation in the presence of the double-stranded RNA-specific ribonuclease RNase III and the Cas9 protein^{29, 51}. Additionally, Cas9 is thought to be the sole protein responsible for the crRNA-guided silencing of foreign DNA^{6, 43, 137}.

The type II CRISPR-Cas, specifically, with the Cas9 protein, is an overall easier, cheaper, and more efficient gene editing tool compared to the pre-existing editing technology, such as ZFNs and TALENs³¹.

Mechanism for CRISPR-Cas 9

The CRISPR-Cas9 system-one of the most prominent and widely used gene editing tools-has the same three main stages of a CRISPR-Cas system: adaptation, expression, and interference.

Adaptation

For the adaptation stage, a complex of Cas proteins may encounter a PAM in the foreign/invading DNA, where the complex will induce 2 double strand breaks close to it; The resulting fragment of foreign DNA (Termed protospacer, is the target) is then integrated into the CRISPR locus between 2 repeats and becomes a spacer around the proximal end of the CRISPR array^{31, 158, 170}.

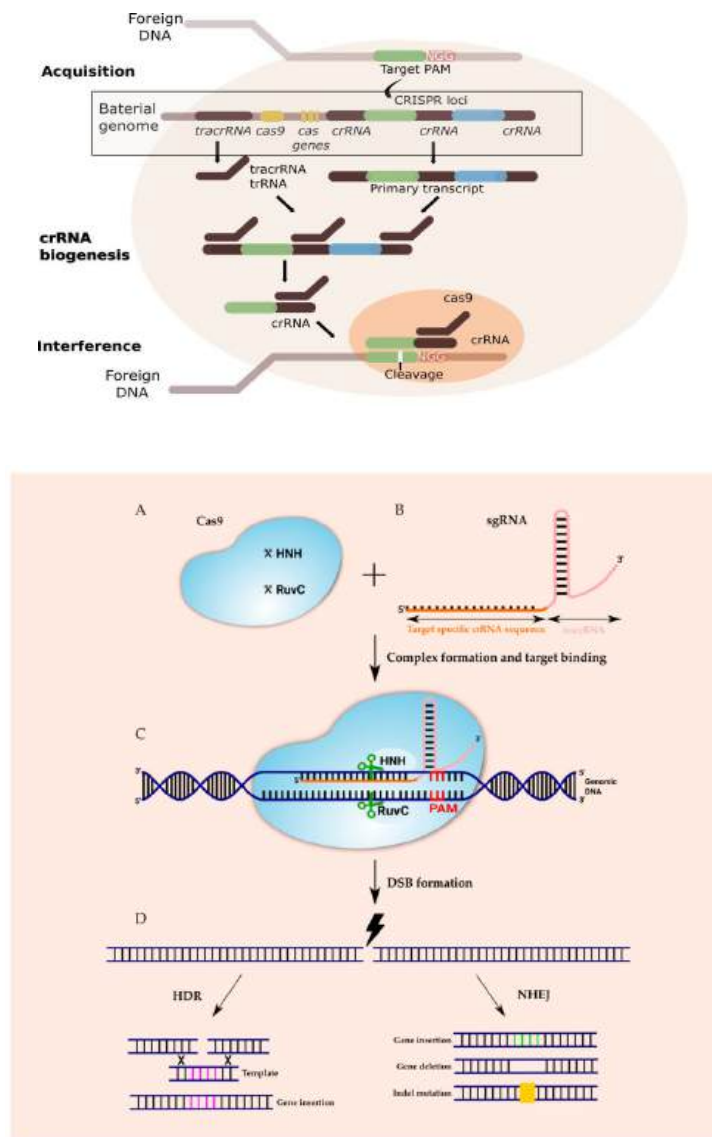


Fig 7-8: The process of CRISPR-Cas9 DNA editing

Expression

For the expression stage, two main events occur in the genome.

One event that occurs is the expression of the Cas9 gene that results in the Cas9 enzyme, a DNA endonuclease that consists of two nuclease domains—the RuvC domain, which cleaves non-complementary DNA strands, and the HNH domain, which cleaves complementary DNA strands—that are responsible for inducing a double-strand break in the target DNA sequence³¹.

The other event—which occurs within the CRISPR Locus—is the transcription of the repeat-spacer element derived from the protospacer, which yields pre-crRNA (**precursor-CRISPR RNA**). Additionally, transactivating crRNAs (tracrRNA) are also produced from another section in the CRISPR locus.

The tracrRNA anneals with the pre-crRNA and directs the maturation of the pre-crRNA into mature crRNA in the presence of the endoribonuclease RNase III and Cas9(also called Csn1) protein^{17, 29, 64}. The resulting crRNA is also reduced to a length of 20 nucleotides(nt) through trimming of the 5' end by other nucleases⁶⁴.

This results in a single-guide RNA (sgRNA) composed of two parts: the crRNA, whose 20 nt 5' end sequence is complementary to the protospacer in the target DNA, and the tracrRNA that's connected to the crRNA which will help bind the entire sgRNA to the Cas9 protein³¹.

Overall, the result of these 2 parallel processes is that a CRISPR–Cas9/sgRNA genome editing complex capable of searching for its target is formed, with the sgRNA binding the complex to the target DNA, and the Cas9 protein actually editing the DNA⁶³.

Interference

Ultimately, the interference stage consists of the Cas9/sgRNA searching for its target, with its search and recognition process requiring both a complementary nt base pairing between the 20-nt crRNA and a protospacer in the target DNA, as well as the presence of the Cas9's corresponding PAM sequence that's close to the target site^{44, 66}.

The PAM sequence is crucial for the complex's ability to differentiate between the CRISPR locus itself and foreign sequences, preventing autoimmunity¹⁰⁵, and even single mutations in the PAM can disable Cas9's DNA cleaving ability in vitro while allowing foreign DNA to avoid the complex^{8, 65, 66}. The PAM sequence that the most common Cas9 type SpCas9(from *Streptococcus pyogenes*) corresponds with is 5'-NGG-3', with N being any 1 of the 4 DNA nucleotides^{24, 59}.

Cas9 initiates the target DNA search for the complex by first probing for its corresponding PAM sequence before interrogating the DNA adjacent to the PAM to check for potential guide-RNA complementarity¹⁴⁷. The target recognition itself occurs through 3-dimensional collisions where Cas9 rapidly splits from the DNA that doesn't contain the appropriate PAM sequence, and the dwell time depends on the complementarity between guide RNA and adjacent DNA when the proper PAM sequence is present^{77, 101, 147}.

Once Cas9 has found the target site with the appropriate PAM sequence, it triggers local DNA denaturing, which causes the DNA strand to split in 2 at the PAM-adjacent target site; this is followed by the crRNA binding to the complementary target strand to form a RNA-DNA hybrid alongside the displaced noncomplementary DNA strand—which is all termed as the R-loop—from the PAM-proximal to PAM-distal ends^{147, 153}.

After the R-loop is formed, the Cas9 is now activated for DNA cleavage¹⁴⁷. The Cas9's two nuclease domains each cut one strand of the target three bp from the PAM sequence to produce a mainly blunt double strand break(DSB)^{44, 66}.

After this, two possible DNA repair mechanisms can occur as a result of the DSB from the Cas9 enzyme: non-homologous end joining (NHEJ) and homology-directed repair (HDR)^{54, 94}.

NHEJ consists of the repair of the DSB by joining DNA fragments through an enzymatic process in the absence of exogenous homologous DNA. This process is active in all phases of the

cell cycle. It is a predominant and efficient cellular repair mechanism that is most active in cells, but it is error-prone and may result in the random insertion/deletion (indels) of base pairs at the cleavage site, which can lead to the generation of frameshift mutation or premature stop codon^{92, 115, 133, 172}.

In contrast, HDR is a highly precise repair mechanism that requires the use of a homologous DNA template and is the most active in the late S and G2 phases of the cell cycle. In CRISPR gene editing, HDR requires a large amount of homologous donor (exogenous) DNA templates from either sister chromatids or foreign DNA, which is used in precise insertions and base substitutions at the DSB site^{13, 90, 94}.

Applications for CRISPR-Cas9

In general, the CRISPR-Cas9 system could be used for many applications.

Functional Genome Screening

One application of CRISPR-Cas9 is in genome functional screening to identify key genes involved in various biological processes across different biological models. It's being used in genome-wide, targeted loss-of-function screens as an alternative screening system to RNA interference (RNAi)³⁹. An instance of this is when a genome-scale CRISPR–Cas9 knockout (GeCKO) library was used to identify genes essential for cell viability in cancer and pluripotent stem cells and was then used to screen for genes whose loss is involved in resistance to vemurafenib (treatment for late-stage melanoma) in a melanoma model¹⁴³. Additionally, other studies have identified host genes important for the intoxication of cells by diphtheria and anthrax toxins using the GeCKO library, which was supported by functional validation¹⁷⁵.

Overall, the application of CRISPR screening in functional genomics is helping researchers to discover new gene functions and, in the future, may help explore the molecular mechanisms of a variety of cellular functions³¹.

Virus Treatment

Another application is related to its potential use as a therapy for chronic viral infections⁸⁵. Most antiviral therapies for conditions such as human immunodeficiency virus(HIV) and hepatitis B virus (HBV) aren't a clinical cure because of the inability to remove the viral genome from the infected host cell due to a latent state where viruses minimize their activity inside the host cell to avoid host immune surveillance. As a result, the latency-related life cycles of these viruses play a key role in the incurability of chronic infections, and patients with the virus will have to take antiviral drugs for the rest of their lives.

But CRISPR itself has shown great promise for the treatment of different viral infections in experiments: one instance is in a study where CRISPR–Cas9 has been shown to inhibit HBV replication and gene expression with the inhibition sustained for 3 days after the CRISPR–Cas9 administration, and the results suggest that the Cas system may provide a simple, inexpensive and short-term process for mammalian genome modification¹⁷⁴; CRISPR–Cas9 delivered by

lentivirus significantly decreasing HIV-1 replication in infected primary CD4+ T cell cultures and reducing the viral load in ex vivo CD4+ T cell culture obtained from HIV-1 infected patients suggests that CRISPR-Cas9 can act as a new type of therapy for HIV, and CRISPR-Cas9 may be used as a novel and efficient platform for the cure of AIDS itself⁷⁰; CRISPR-cas9 when used on a Burkitt's lymphoma cell line with latent Epstein-Barr virus (EBV) infection was able to get rid of EBV completely for 25% of the cells, and caused significant EBV load decrease for another 50% of the treated cells when compared to the untreated ones¹⁶².

Overall, the use of CRISPR-Cas9 as an antiviral strategy is a promising prospect, with various in vitro and in vivo studies, giving rise to potential clinical applications for humans that could result in treatment or prevention methods for pathogenic viral infections in humans.

Combating Antibiotic-Resistant Bacteria

CRISPRs can also be used to control and combat antibiotic-resistant bacteria.

One experiment where 2 CRISPR-dCas9 systems were used to repress antibiotic resistance in MRSA (Methicillin-resistant *Staphylococcus aureus*-a human pathogen which is resistant to β -lactam antibiotics such as penicillin) by targeting the *mecA* methicillin resistance gene resulted in a 77% decrease in gene expression in the CRISPR-treated samples, but it did not kill the bacteria, which makes it a viable option for treatment against antibiotic-resistant bacteria¹⁶³.

Another experiment with MRSA introduced a nonviral delivery method (Known as Cr-Nanocomplex) for CRISPR targeting *mecA* (major antibiotic resistance gene involved in MRSA,) which consisted of SpCas9 covalently modified with a cationic polymer called polyethyleneimine (bPEI), as the carrier for packaging sgRNA and boosting their delivery to bacteria. Furthermore, the cultured MRSA strains which were treated with the Cr-Nanocomplex were not able to grow in agar media, including oxacillin at the dose 6 $\mu\text{g/mL}$, while the untreated samples could grow^{71, 164}.

Overall, the use of CRISPR-Cas9 to specifically remove resistance genes in bacteria can be a powerful tool to counteract antibiotic resistance and could be a part of the solution to keep antibiotics working; CRISPR-Cas9 would also have to be used as a conventional broad-spectrum antibiotic, which can potentially target only antibiotic-resistant bacteria while leaving the rest alone³¹.

Cancer Therapy

Another application is CRISPR as a potential cancer treatment³¹. This is shown through multiple studies and experiments. One is in an experiment where epidermal growth factor receptor (EGFR) mutations in EGFR mutation-bearing tumors in vivo were targeted by CRISPR-Cas9 through adenovirus vector delivery which had disrupted the EGFR mutation (L858R) and resulted in cancer cell death and significant reduction of tumor size⁸⁰.

Another is in a study where CRISPR-Cas9 was used to generate chimeric antigen receptor T (CAR-T) cells, which can recognize specific antigens on cancer cells. These CAR-T

cells were used in therapy for patients with acute lymphoblastic leukemia which targeted CD19 and resulted in an effective treatment for the lymphoblastic leukemia while being related to a high remission rate for patients where Stem-cell transplantation was ineffective¹⁰⁶. Some problems face CRISPR cancer therapy, however. One of these problems is the lack of a safe and efficient delivery mechanism that can be applied in clinical trials, alongside the fact that tumor heterogeneity caused by tumors normally having different subclones with different DNA mutations within can make creating the treatment difficult with CRISPR^{21, 46}.

Another problem is editing efficiency, as the effectiveness of DSB repair with HDR and NHEJ can vary in different cell types and states, and the potentially low editing efficiency caused by that factor in certain cells can cause a lack of effectiveness from the treatment^{25, 173}.

Additionally, Cas9—due to being a protein of bacterial origin—can activate the body's immune system, again lowering the effectiveness of the treatment²².

But despite all the problems, CRISPR is still crucial in cancer research³¹; If its safety, specificity, and efficacy is optimized enough for clinical use, then it could be a widespread cancer treatment.

CRISPR Base Editing

Base editing is a form of genome editing that enables the direct and irreversible conversion of one basepair to another at a target genomic locus without requiring DSBs, or HDR processes with their necessary donor DNA templates^{78, 79, 117}. Compared to the standard genome editing methods to introduce point mutations, base editing can do so more efficiently⁷⁹, with far less problems such as stochastic insertions/deletions (indels) or translocations^{76, 78, 79, 117, 130, 184}.

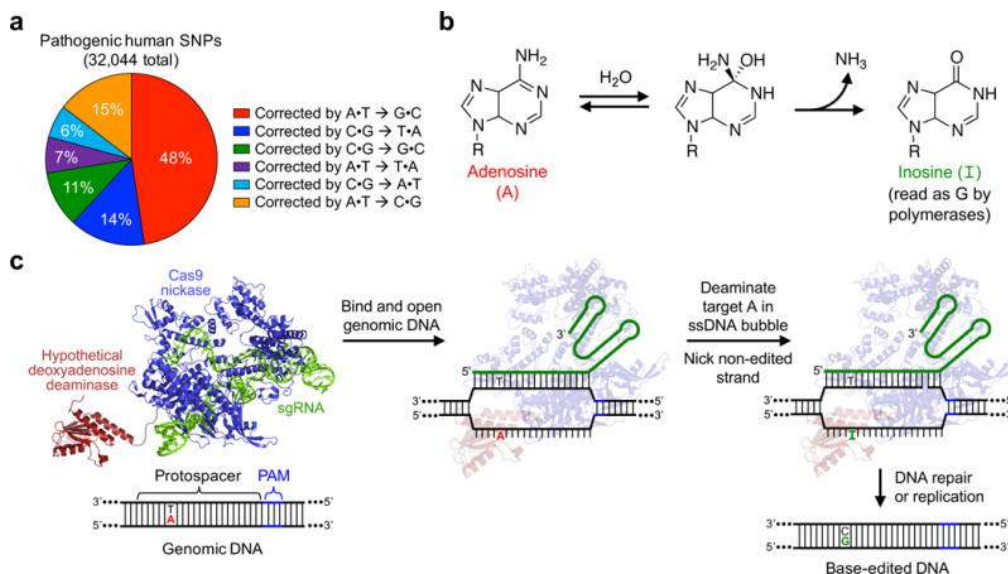


Fig 9: In b, the deamination of Adenosine to Inosine, which is read as G by polymerases. In c, the general structure of how an ABE would work with a deoxyadenosine deaminase working alongside an enzymically impaired Cas9, alongside sgRNA to find a spot in the target DNA and catalyze the deamination of Adenosine to Inosine⁴⁵

There are two main categories of Base editors(BE): cytosine base editors (CBEs) and adenine base editors (ABEs), which are all used with sgRNA for guidance³¹.

Regarding CBEs, one of the most commonly used versions are 3rd generation designs (CBE3), which consist of four main parts:

The first piece is a catalytically impaired CRISPR-Cas9 mutant that is unable to create DSBs (known as nuclease-dead Cas9 or dCas9 for short), the second is a single-strand-specific enzyme capable of converting one base to another with an example in this case being a cytidine deaminase that converts Cytosine(C) to Uracil(U) within a 5-nt window in a single-stranded DNA containing bubble created by the Cas9, 3rd is a uracil glycosylase inhibitor (UGI) that impedes uracil removal and the downstream processes that decrease the base editing efficiency and product purity⁷⁸, and the 4th is an enzymatic nickase that nicks the other non-edited DNA strand complimentary to the first strand, directing cellular DNA repair processes to replace the G-containing DNA strand^{78, 79}.

With all these components, BE3 can carry out the conversion of C•G to T•A efficiently and permanently in many different organisms, from bacteria and plants^{97, 179}, to things such as mice^{75, 130, 183} and human embryos^{88, 89}.

Base editors have expanded in capability through the development of base editors with different properties: different PAM compatibilities⁷⁶, narrower editing windows⁷⁶, enhanced DNA specificity¹³⁰, and small-molecule dependence¹⁵⁴.

Regarding ABEs, these are a relatively recent development, but they share similarities with CBEs in terms of their mechanical design, except for some major differences. One big difference is that CBEs and ABEs have a different single-strand-specific enzyme capable of converting one base to another, with ABEs having a deoxyadenosine deaminase that can catalyze the deamination of Adenine to Inosine in DNA where the Inosine is interpreted as Guanine^{45, 72}. A note about this is that there were no naturally occurring single-stranded DNA Adenine deaminases; only RNA adenosine deaminases were found, which were all unable to edit DNA⁴⁵.

But it was engineered through the directed evolution of *Escherichia coli*'s tRNA adenosine deaminase(TadA) which was chosen since it shares homology with the APOBEC enzyme (cytidine deaminase) used in the previous CBEs, and some APOBECs bind single-stranded DNA in a conformation that resembles tRNA bound to the TadA¹⁴⁴.

The experiment first consisted of introducing a C•G to T•A mutation in the CamR gene of *Escherichia coli*—a gene that codes for chloramphenicol resistance in bacteria—in the form of an H193Y substitution that would weaken the bacteria's chloramphenicol resistance⁴⁵. Then, the researchers had created unbiased libraries of *Escherichia coli*(ec)/TadA-dCas9 fusions that contained mutations only in the adenine deaminase portion of the construct to avoid altering the favorable properties of the Cas9 portion of the editor, and were then transferred into the *E. coli* harboring the CamR H193Y mutation with the idea that A•T to G•C conversion at the H193Y mutation should restore chloramphenicol resistance, linking the TadA's Adenine base editing activity to bacterial survival⁴⁵.

The surviving bacterial colonies that were able to catalyze the A•T to G•C in the CamR gene had the TadA mutations, A106V and D108N⁴⁵. These mutated TadAs were then incorporated into a mammalian codon-optimized TadA–Cas9 nickase fusion construct, where the dCas9 was replaced with the Cas9 D10A nickase used in BE3 to manipulate cellular DNA repair to favor the desired base editing outcomes⁷⁹ alongside a C-terminal nuclear localization signal (NLS)⁴⁵. The resulting TadA*–XTEN–nCas9–NLS construct—where TadA* is the new TadA variant and XTEN is a 16-amino acid linker used in BE33— was designated as ABE1.2, which was capable of carrying out very low efficiency, but observable A•T to G•C editing⁴⁵.

From here on out, efforts would be made to optimize ABEs through engineering a single-chain heterodimer, consisting of a wild-type non-catalytic TadA monomer and an evolved TadA monomer(TadA-TadA*). And further work would result in prominent versions such as ABE7.10 and ABE8e that would convert A•T to G•C base pairs efficiently with very high product purity and very low rates of indels^{45, 72}.

Additionally, ABEs compared with CBEs yield a far cleaner product with fewer mistakes and reports of significant off-target (A-to-non-G) edits to date⁷². Adding on to this, unlike CBEs with their UGI mediated inhibition of Uracil-DNA glycosylase(UDG-enzyme that removes Uracil), ABE editing in cells that lacked alkyl adenine DNA glycosylase (AAG)—the enzyme known to recognize and remove inosine in DNA which would in theory inhibit ABE ability—failed to increase the editing efficiency or product purity compared with cells containing AAG, which shows the sheer converting power of ABE compared to CBE⁷².

CRISPR's Use for Curing Chronic Disease

CRISPR-Cas9 base editing could be used to cure chronic genetic diseases, such as Wilson's Disease, that would otherwise be incurable and require treatment for the rest of the patient's lifetime. This would be done by first identifying the specific mutation in the target gene (ATP7B in this case) that results in the disease. Mutations can occur in multiple different places on a gene, such as H1069Q and R778L in patients with Wilson's Disease, making genomic screening a valuable and necessary precursor to a newly designed treatment. The mutation itself can be of types that are in different places within the gene for each person which would require the use of genomic screening, with Wilson's Disease having many different missense mutations, such as H1069Q and R778L. Once the mutation is identified, the sequence of the mutation area would be inserted into a CRISPR locus, where the proper sgRNA would be generated alongside Cas9 with an engineered base editor; this would then form the basis for a Cas9/sgrNA genome editing complex. The complex would then be injected into the patient in a specific bodily region most affected by the disease. For Wilson's Disease, a medical team would most likely target the liver for gene editing complex delivery. Finally, the patient would be monitored to ensure there are no negative effects from treatment.

The most apparent benefit is that CRISPR-Cas9 base editing would possibly be able to cure the genetic disease for good; moreover, another benefit is that with a one-time treatment, the

patient receiving it would never have to spend additional money on temporary treatments, saving thousands of dollars a year.

The idea of CRISPR curing a genetic disease has been made a reality quite recently. A real-life example of CRISPR-based editing used to treat disease is the case of an infant named “KJ” who received treatment for a rare inherited metabolic disease in 2025, detailed in a study by The New England Journal of Medicine. Specifically, KJ experienced a severe carbamoyl phosphate synthetase 1 (CPS1) deficiency. This is a rare genetic disease that causes the lack of a liver enzyme in the liver needed to convert ammonia (generated from the normal breakdown of proteins in the body to urea), which causes ammonia to build up to toxic levels in the subject and can damage organs such as the brain and liver. Patients with CPS1 deficiency, like KJ would normally be treated with a liver transplant; however, to undergo a liver transplant, patients would need to be medically stable and old enough to handle such a major procedure—neither of which was true for KJ.

Ahrens-Nicklas and Kiran Musunuru MD, PhD (Barry J. Gertz Professor for Translational Research in Penn’s Perelman School of Medicine) are co-corresponding authors on the published report and began collaborating in 2023 to study the possibility of creating customized gene editing therapies for rare metabolic disorders. They had mainly focused on urea cycle disorders. Soon after KJ’s birth, they were able to identify KJ’s specific variant of CPS1 through genetic sequencing. The team, including Nicklas and Musunuru, was able to design and manufacture a CRISPR-based editing therapy in 6 months. In late February 2025, the treatment was delivered to KJ’s liver via nanoparticles to correct KJ’s lack of functional CPS1 enzyme, and then he received follow-up doses in March and April of 2025. As of the time of reporting, KJ has experienced no serious side effects from treatment; additionally, KJ has been able to increase consumption of protein and has been weaned off of nitrogen scavenger medication. KJ has also demonstrated recovery from certain typical childhood illnesses like rhinovirus (virus that’s predominant cause of common cold) without ammonia building up in his body. Longer follow-up is needed to fully evaluate the benefits of the therapy, but so far the treatment seems promising¹¹⁴.

Although KJ will need to be monitored carefully for the rest of his life due to possible immunogenic or genetic side effects, his case demonstrates the possibility for CRISPR to be able to treat genomic diseases that would otherwise put great strain and pressure on thousands of people.

Acknowledgements

Thanks to Dr. Daniel Plymire and Theodore Press for providing assistance in the form of helpful discussions and inspiration for the paper.

Works Cited

1. Ahmad, Ahsan, et al. "Liver Transplantation for Wilson Disease." *Handbook of Clinical Neurology*, vol. 142, 2017, pp. 193–204. *PubMed*, <https://doi.org/10.1016/B978-0-444-63625-6.00016-1>.
2. Andrés: *Impact of Reactive Species on Amino Acids—Biol...* - Google Scholar. https://scholar.google.com/scholar_lookup?title=Impact%20of%20reactive%20species%20on%20amino%20acids-biological%20relevance%20in%20proteins%20and%20induced%20pathologies&publication_year=2022&author=C.M.C.%20Andr%C3%A9s&author=J.M.%20P%C3%A9rez%20de%20la%20Lastra&author=C.%20Andr%C3%A9s%20Juan&author=F.J.%20Plou&author=E.%20P%C3%A9rez-Lebe%C3%B1a. Accessed 27 July 2025.
3. Avan, Abolfazl, et al. "The Role of Zinc in the Treatment of Wilson's Disease." *International Journal of Molecular Sciences*, vol. 23, no. 16, Jan. 2022, p. 9316. *www.mdpi.com*, <https://doi.org/10.3390/ijms23169316>.
4. Banci, Lucia, et al. "Human Superoxide Dismutase 1 (hSOD1) Maturation through Interaction with Human Copper Chaperone for SOD1 (hCCS)." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 34, Aug. 2012, pp. 13555–60. *PubMed*, <https://doi.org/10.1073/pnas.1207493109>.
5. Bandmann, Oliver, et al. "Wilson's Disease and Other Neurological Copper Disorders." *The Lancet. Neurology*, vol. 14, no. 1, Jan. 2015, pp. 103–13. *PubMed Central*, [https://doi.org/10.1016/S1474-4422\(14\)70190-5](https://doi.org/10.1016/S1474-4422(14)70190-5).
6. Barrangou, Rodolphe, et al. "CRISPR Provides Acquired Resistance against Viruses in Prokaryotes." *Science (New York, N.Y.)*, vol. 315, no. 5819, Mar. 2007, pp. 1709–12. *PubMed*, <https://doi.org/10.1126/science.1138140>.
7. Bhaya, Devaki, et al. "CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation." *Annual Review of Genetics*, vol. 45, 2011, pp. 273–97. *PubMed*, <https://doi.org/10.1146/annurev-genet-110410-132430>.
8. Bikard, David, et al. "CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In Vivo Bacterial Infection." *Cell Host & Microbe*, vol. 12, no. 2, Aug. 2012, pp. 177–86. *www.cell.com*, <https://doi.org/10.1016/j.chom.2012.06.003>.
9. Bjørø, K., et al. "Secondary Osteoporosis in Liver Transplant Recipients: A Longitudinal Study in Patients with and without Cholestatic Liver Disease." *Scandinavian Journal of Gastroenterology*, vol. 38, no. 3, Mar. 2003, pp. 320–27.
10. Boga, Salih, et al. "Hepatic Features of Wilson Disease." *Handbook of Clinical Neurology*, vol. 142, 2017, pp. 91–99. *PubMed*, <https://doi.org/10.1016/B978-0-444-63625-6.00009-4>.
11. Bolotin, Alexander, et al. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin." *Microbiology (Reading, England)*, vol. 151, no. Pt 8, Aug. 2005, pp. 2551–61. *PubMed*, <https://doi.org/10.1099/mic.0.28048-0>.

12. Borchard, Sabine, et al. "The Exceptional Sensitivity of Brain Mitochondria to Copper." *Toxicology in Vitro*, vol. 51, Sept. 2018, pp. 11–22. *ScienceDirect*, <https://doi.org/10.1016/j.tiv.2018.04.012>.
13. Branzei, Dana, and Marco Foiani. "Regulation of DNA Repair throughout the Cell Cycle." *Nature Reviews. Molecular Cell Biology*, vol. 9, no. 4, Apr. 2008, pp. 297–308. *PubMed*, <https://doi.org/10.1038/nrm2351>.
14. Brewer, George J., et al. "Treatment of Wilson Disease with Ammonium Tetrathiomolybdate: IV. Comparison of Tetrathiomolybdate and Trientine in a Double-Blind Study of Treatment of the Neurologic Presentation of Wilson Disease." *Archives of Neurology*, vol. 63, no. 4, Apr. 2006, pp. 521–27. *PubMed*, <https://doi.org/10.1001/archneur.63.4.521>.
15. Broniek-Kowalik, Karina, et al. "Anterior Segment Optical Coherence Tomography (AS-OCT) as a New Method of Detecting Copper Deposits Forming the Kayser-Fleischer Ring in Patients with Wilson Disease." *Acta Ophthalmologica*, vol. 97, no. 5, Aug. 2019, pp. e757–60. *PubMed*, <https://doi.org/10.1111/aos.14009>.
16. Buksińska-Lisik, Małgorzata, et al. "Cardiac Assessment in Wilson's Disease Patients Based on Electrocardiography and Echocardiography Examination." *Archives of Medical Science : AMS*, vol. 15, no. 4, July 2019, pp. 857–64. *PubMed Central*, <https://doi.org/10.5114/aoms.2017.69728>.
17. C, Liao, and Beisel Cl. "The tracrRNA in CRISPR Biology and Technologies." *Annual Review of Genetics*, vol. 55, Nov. 2021. *pubmed.ncbi.nlm.nih.gov*, <https://doi.org/10.1146/annurev-genet-071719-022559>.
18. Camarata, Michelle A., et al. "Zinc Maintenance Therapy for Wilson Disease: A Comparison Between Zinc Acetate and Alternative Zinc Preparations." *Hepatology Communications*, vol. 3, no. 8, July 2019, pp. 1151–58. *PubMed Central*, <https://doi.org/10.1002/hep4.1384>.
19. Carta, Mauro Giovanni, et al. "Bipolar Disorders and Wilson's Disease." *BMC Psychiatry*, vol. 12, May 2012, p. 52. *PubMed*, <https://doi.org/10.1186/1471-244X-12-52>.
20. Cater, Michael A., et al. "Intracellular Trafficking of the Human Wilson Protein: The Role of the Six N-Terminal Metal-Binding Sites." *The Biochemical Journal*, vol. 380, no. Pt 3, June 2004, pp. 805–13. *PubMed*, <https://doi.org/10.1042/BJ20031804>.
21. Chen, Minjiang, et al. "CRISPR-Cas9 for Cancer Therapy: Opportunities and Challenges." *Cancer Letters*, vol. 447, Apr. 2019, pp. 48–55. *PubMed*, <https://doi.org/10.1016/j.canlet.2019.01.017>.
22. Cheng, Xing, et al. "CRISPR/Cas9 for Cancer Treatment: Technology, Clinical Applications and Challenges." *Briefings in Functional Genomics*, vol. 19, no. 3, May 2020, pp. 209–14. *PubMed*, <https://doi.org/10.1093/bfpg/ela001>.
23. Cong, Le, et al. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science (New York, N.Y.)*, vol. 339, no. 6121, Feb. 2013, pp. 819–23. *PubMed*, <https://doi.org/10.1126/science.1231143>.

24. Cong, Le, and Feng Zhang. "Genome Engineering Using CRISPR-Cas9 System." *Methods in Molecular Biology (Clifton, N.J.)*, vol. 1239, 2015, pp. 197–217. *PubMed*, https://doi.org/10.1007/978-1-4939-1862-1_10.
25. Cox, David Benjamin Turitz, et al. "Therapeutic Genome Editing: Prospects and Challenges." *Nature Medicine*, vol. 21, no. 2, Feb. 2015, pp. 121–31. *PubMed Central*, <https://doi.org/10.1038/nm.3793>.
26. Culotta, Valeria Cizewski, et al. "Activation of Superoxide Dismutases: Putting the Metal to the Pedal." *Biochimica Et Biophysica Acta*, vol. 1763, no. 7, July 2006, pp. 747–58. *PubMed*, <https://doi.org/10.1016/j.bbamcr.2006.05.003>.
27. Czlonkowska, Anna, et al. "Nature Reviews Disease Primers Article: Wilson Disease." *Nature Reviews. Disease Primers*, vol. 4, no. 1, Sept. 2018, p. 21. *PubMed Central*, <https://doi.org/10.1038/s41572-018-0018-3>.
28. Czlonkowska, Anna, and Tomasz Litwin. "Wilson Disease - Currently Used Anticopper Therapy." *Handbook of Clinical Neurology*, vol. 142, 2017, pp. 181–91. *PubMed*, <https://doi.org/10.1016/B978-0-444-63625-6.00015-X>.
29. Deltcheva, Elitza, et al. "CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III." *Nature*, vol. 471, no. 7340, Mar. 2011, pp. 602–07. *PubMed*, <https://doi.org/10.1038/nature09886>.
30. Dhawan, Anil, et al. "Wilson's Disease in Children: 37-Year Experience and Revised King's Score for Liver Transplantation." *Liver Transplantation: Official Publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society*, vol. 11, no. 4, Apr. 2005, pp. 441–48. *PubMed*, <https://doi.org/10.1002/lt.20352>.
31. E, Janik, et al. "Various Aspects of a Gene Editing System-CRISPR-Cas9." *International Journal of Molecular Sciences*, vol. 21, no. 24, Dec. 2020. *pubmed.ncbi.nlm.nih.gov*, <https://doi.org/10.3390/ijms21249604>.
32. European Association for Study of Liver. "EASL Clinical Practice Guidelines: Wilson's Disease." *Journal of Hepatology*, vol. 56, no. 3, Mar. 2012, pp. 671–85. *PubMed*, <https://doi.org/10.1016/j.jhep.2011.11.007>.
33. Fatemi, Negah, and Bibudhendra Sarkar. "Molecular Mechanism of Copper Transport in Wilson Disease." *Environmental Health Perspectives*, 110 Suppl 5, no. Suppl 5, Oct. 2002, pp. 695–98. *PubMed*, <https://doi.org/10.1289/ehp.02110s5695>.
34. Fenu, M., et al. "Kayser-Fleischer Ring in Wilson's Disease: A Cohort Study." *European Journal of Internal Medicine*, vol. 23, no. 6, Sept. 2012, pp. e150-156. *PubMed*, <https://doi.org/10.1016/j.ejim.2012.04.005>.
35. Ferenci, Peter, et al. "Diagnosis and Phenotypic Classification of Wilson Disease." *Liver International: Official Journal of the International Association for the Study of the Liver*, vol. 23, no. 3, June 2003, pp. 139–42. *PubMed*, <https://doi.org/10.1034/j.1600-0676.2003.00824.x>.
36. Ferenci. "Regional Distribution of Mutations of the ATP7B Gene in Patients with Wilson

- Disease: Impact on Genetic Testing.” *Human Genetics*, vol. 120, no. 2, Sept. 2006, pp. 151–59. *PubMed*, <https://doi.org/10.1007/s00439-006-0202-5>.
37. Ferguson-Miller, Shelagh, and Gerald T. Babcock. “Heme/Copper Terminal Oxidases.” *Chemical Reviews*, vol. 96, no. 7, Nov. 1996, pp. 2889–908. *PubMed*, <https://doi.org/10.1021/cr950051s>.
 38. Festa, Richard A., and Dennis J. Thiele. “Copper: An Essential Metal in Biology.” *Current Biology*, vol. 21, no. 21, Nov. 2011, pp. R877–83. *www.cell.com*, <https://doi.org/10.1016/j.cub.2011.09.040>.
 39. Ford, Kyle, et al. “Functional Genomics via CRISPR-Cas.” *Journal of Molecular Biology*, vol. 431, no. 1, Jan. 2019, pp. 48–65. *PubMed Central*, <https://doi.org/10.1016/j.jmb.2018.06.034>.
 40. Forman, S. J., et al. “Hemolytic Anemia in Wilson Disease: Clinical Findings and Biochemical Mechanisms.” *American Journal of Hematology*, vol. 9, no. 3, 1980, pp. 269–75. *PubMed*, <https://doi.org/10.1002/ajh.2830090305>.
 41. Fridovich, Irwin. “Superoxide Radical and Superoxide Dismutases.” *Oxygen and Living Processes*, edited by Daniel L. Gilbert, Springer, 1981, pp. 250–72. *Springer Link*, https://doi.org/10.1007/978-1-4612-5890-2_13.
 42. García, Noemí, et al. “Copper Induces Permeability Transition through Its Interaction with the Adenine Nucleotide Translocase.” *Cell Biology International*, vol. 31, no. 9, Sept. 2007, pp. 893–99. *ScienceDirect*, <https://doi.org/10.1016/j.cellbi.2007.02.003>.
 43. Garneau, Josiane E., et al. “The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA.” *Nature*, vol. 468, no. 7320, Nov. 2010, pp. 67–71. *PubMed*, <https://doi.org/10.1038/nature09523>.
 44. Gasiunas, Giedrius, et al. “Cas9–crRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria.” *Proceedings of the National Academy of Sciences*, vol. 109, no. 39, Sept. 2012, pp. E2579–86. *pnas.org (Atypon)*, <https://doi.org/10.1073/pnas.1208507109>.
 45. Gaudelli, Nicole M., et al. “Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage.” *Nature*, vol. 551, no. 7681, Nov. 2017, pp. 464–71. *PubMed Central*, <https://doi.org/10.1038/nature24644>.
 46. Gerlinger, Marco, et al. “Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing.” *The New England Journal of Medicine*, vol. 366, no. 10, Mar. 2012, pp. 883–92. *PubMed Central*, <https://doi.org/10.1056/NEJMoa1113205>.
 47. Ghosh, Upasana, et al. “Challenges and Dilemmas in Pediatric Hepatic Wilson’s Disease.” *World Journal of Hepatology*, vol. 15, no. 10, Oct. 2023, pp. 1109–26. *PubMed*, <https://doi.org/10.4254/wjh.v15.i10.1109>.
 48. Girotti, Albert W. “Mechanisms of Lipid Peroxidation.” *Journal of Free Radicals in Biology & Medicine*, vol. 1, no. 2, Jan. 1985, pp. 87–95. *ScienceDirect*, [https://doi.org/10.1016/0748-5514\(85\)90011-X](https://doi.org/10.1016/0748-5514(85)90011-X).
 49. Godde, James S., and Amanda Bickerton. “The Repetitive DNA Elements Called

- CRISPRs and Their Associated Genes: Evidence of Horizontal Transfer among Prokaryotes.” *Journal of Molecular Evolution*, vol. 62, no. 6, June 2006, pp. 718–29. *PubMed*, <https://doi.org/10.1007/s00239-005-0223-z>.
50. Gonwa, T. A., et al. “End-Stage Renal Disease (ESRD) after Orthotopic Liver Transplantation (OLT) Using Calcineurin-Based Immunotherapy: Risk of Development and Treatment.” *Transplantation*, vol. 72, no. 12, Dec. 2001, pp. 1934–39. *PubMed*, <https://doi.org/10.1097/00007890-200112270-00012>.
 51. Gottesman, Susan. “Microbiology: Dicing Defence in Bacteria.” *Nature*, vol. 471, no. 7340, Mar. 2011, pp. 588–89. *PubMed*, <https://doi.org/10.1038/471588a>.
 52. Grissa, Ibtissem, et al. “The CRISPRdb Database and Tools to Display CRISPRs and to Generate Dictionaries of Spacers and Repeats.” *BMC Bioinformatics*, vol. 8, May 2007, p. 172. *PubMed Central*, <https://doi.org/10.1186/1471-2105-8-172>.
 53. Gwirtsman, H. E., et al. “Case Report of Anorexia Nervosa Associated with Wilson’s Disease.” *The International Journal of Eating Disorders*, vol. 13, no. 2, Mar. 1993, pp. 241–44. *PubMed*, [https://doi.org/10.1002/1098-108x\(199303\)13:2%253C241::aid-eat2260130213%253E3.0.co;2-y](https://doi.org/10.1002/1098-108x(199303)13:2%253C241::aid-eat2260130213%253E3.0.co;2-y).
 54. Haapaniemi, Emma, et al. “CRISPR-Cas9 Genome Editing Induces a P53-Mediated DNA Damage Response.” *Nature Medicine*, vol. 24, no. 7, July 2018, pp. 927–30. *PubMed*, <https://doi.org/10.1038/s41591-018-0049-z>.
 55. Habka, Dany, et al. “Future Economics of Liver Transplantation: A 20-Year Cost Modeling Forecast and the Prospect of Bioengineering Autologous Liver Grafts.” *PloS One*, vol. 10, no. 7, 2015, p. e0131764. *PubMed*, <https://doi.org/10.1371/journal.pone.0131764>.
 56. Haft, Daniel H, et al. “A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes.” *PLoS Computational Biology*, vol. 1, no. 6, Nov. 2005, p. e60. *PubMed Central*, <https://doi.org/10.1371/journal.pcbi.0010060>.
 57. Halliwell, Barry, and John M. C. Gutteridge. “[1] Role of Free Radicals and Catalytic Metal Ions in Human Disease: An Overview.” *Methods in Enzymology*, vol. 186, Academic Press, 1990, pp. 1–85. *Oxygen Radicals in Biological Systems Part B: Oxygen Radicals and Antioxidants. ScienceDirect*, [https://doi.org/10.1016/0076-6879\(90\)86093-B](https://doi.org/10.1016/0076-6879(90)86093-B).
 58. Halliwell: *Hydroxyl Radical Is a Significant Player...* - *Google Scholar*. https://scholar.google.com/scholar_lookup?title=Hydroxyl%20radical%20is%20a%20significant%20player%20in%20oxidative%20DNA%20damage%20in%20vivo&publication_year=2021&author=B.%20Halliwell&author=A.%20Adhikary&author=M.%20Dingfelder&author=M.%20Dizdaroglu. Accessed 27 July 2025.
 59. Hirano, Seiichi, et al. “Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9.” *Molecular Cell*, vol. 61, no. 6, Mar. 2016, pp. 886–94. *PubMed*,

- <https://doi.org/10.1016/j.molcel.2016.02.018>.
60. Ishino, Y, et al. "Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product." *Journal of Bacteriology*, vol. 169, no. 12, Dec. 1987, pp. 5429–33. *journals.asm.org (Atypon)*, <https://doi.org/10.1128/jb.169.12.5429-5433.1987>.
 61. Ishino, Yoshizumi, et al. "History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology." *Journal of Bacteriology*, vol. 200, no. 7, Mar. 2018, p. 10.1128/jb.00580-17. *journals.asm.org (Atypon)*, <https://doi.org/10.1128/jb.00580-17>.
 62. Jansen, Ruud, et al. "Identification of Genes That Are Associated with DNA Repeats in Prokaryotes." *Molecular Microbiology*, vol. 43, no. 6, Mar. 2002, pp. 1565–75. *PubMed*, <https://doi.org/10.1046/j.1365-2958.2002.02839.x>.
 63. Jiang, Fuguo, et al. "A Cas9–Guide RNA Complex Preorganized for Target DNA Recognition." *Science*, vol. 348, no. 6242, June 2015, pp. 1477–81. *science.org (Atypon)*, <https://doi.org/10.1126/science.aab1452>.
 64. Jiang, Fuguo, and Jennifer A. Doudna. "CRISPR–Cas9 Structures and Mechanisms." *Annual Review of Biophysics*, vol. 46, no. Volume 46, 2017, May 2017, pp. 505–29. *www.annualreviews.org*, <https://doi.org/10.1146/annurev-biophys-062215-010822>.
 65. Jiang, Wenyan, et al. "RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems." *Nature Biotechnology*, vol. 31, no. 3, Mar. 2013, pp. 233–39. *www.nature.com*, <https://doi.org/10.1038/nbt.2508>.
 66. Jinek, Martin, Krzysztof Chylinski, et al. "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science*, vol. 337, no. 6096, Aug. 2012, pp. 816–21. *science.org (Atypon)*, <https://doi.org/10.1126/science.1225829>.
 67. Jinek, Martin, Alexandra East, et al. "RNA-Programmed Genome Editing in Human Cells." *eLife*, vol. 2, Jan. 2013, p. e00471. *PubMed Central*, <https://doi.org/10.7554/eLife.00471>.
 68. Kallinikos, Panagiotis, et al. "Corneal Nerve Tortuosity in Diabetic Patients with Neuropathy." *Investigative Ophthalmology & Visual Science*, vol. 45, no. 2, Feb. 2004, pp. 418–22. *PubMed*, <https://doi.org/10.1167/iovs.03-0637>.
 69. Kamath, Patrick S., et al. "The Model for End-Stage Liver Disease (MELD)." *Hepatology (Baltimore, Md.)*, vol. 45, no. 3, Mar. 2007, pp. 797–805. *PubMed*, <https://doi.org/10.1002/hep.21563>.
 70. Kaminski, Rafal, et al. "Elimination of HIV-1 Genomes from Human T-Lymphoid Cells by CRISPR/Cas9 Gene Editing." *Scientific Reports*, vol. 6, Mar. 2016, p. 22555. *PubMed Central*, <https://doi.org/10.1038/srep22555>.
 71. Kang, Yoo Kyung, et al. "Nonviral Genome Editing Based on a Polymer-Derivatized CRISPR Nanocomplex for Targeting Bacterial Pathogens and Antibiotic Resistance." *Bioconjugate Chemistry*, vol. 28, no. 4, Apr. 2017, pp. 957–67. *PubMed*, <https://doi.org/10.1021/acs.bioconjchem.6b00676>.

72. Kantor, Ariel, et al. "CRISPR-Cas9 DNA Base-Editing and Prime-Editing." *International Journal of Molecular Sciences*, vol. 21, no. 17, Aug. 2020, p. 6240. *PubMed Central*, <https://doi.org/10.3390/ijms21176240>.
73. Karginov, Fedor V., and Gregory J. Hannon. "The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea." *Molecular Cell*, vol. 37, no. 1, Jan. 2010, p. 7. *PubMed Central*, <https://doi.org/10.1016/j.molcel.2009.12.033>.
74. Kelly, Claire, and Marinos Pericleous. "Wilson Disease: More than Meets the Eye." *Postgraduate Medical Journal*, vol. 94, no. 1112, June 2018, pp. 335–47. *PubMed*, <https://doi.org/10.1136/postgradmedj-2017-135381>.
75. Kim, Kyoungmi, et al. "Highly Efficient RNA-Guided Base Editing in Mouse Embryos." *Nature Biotechnology*, vol. 35, no. 5, May 2017, pp. 435–37. *PubMed*, <https://doi.org/10.1038/nbt.3816>.
76. Kim, Y. Bill, et al. "Increasing the Genome-Targeting Scope and Precision of Base Editing with Engineered Cas9-Cytidine Deaminase Fusions." *Nature Biotechnology*, vol. 35, no. 4, Apr. 2017, pp. 371–76. *PubMed Central*, <https://doi.org/10.1038/nbt.3803>.
77. Knight, Spencer C., et al. "Dynamics of CRISPR-Cas9 Genome Interrogation in Living Cells." *Science*, vol. 350, no. 6262, Nov. 2015, pp. 823–26. *science.org (Atypon)*, <https://doi.org/10.1126/science.aac6572>.
78. Komor, Alexis C., Kevin T. Zhao, et al. "Improved Base Excision Repair Inhibition and Bacteriophage Mu Gam Protein Yields C:G-to-T:A Base Editors with Higher Efficiency and Product Purity." *Science Advances*, vol. 3, no. 8, Aug. 2017, p. eaao4774. *PubMed Central*, <https://doi.org/10.1126/sciadv.aao4774>.
79. Komor, Alexis C., Yongjoo B. Kim, et al. "Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage." *Nature*, vol. 533, no. 7603, Apr. 2016, pp. 420–24. *PubMed Central*, <https://doi.org/10.1038/nature17946>.
80. Koo, Taeyoung, et al. "Selective Disruption of an Oncogenic Mutant Allele by CRISPR/Cas9 Induces Efficient Tumor Regression." *Nucleic Acids Research*, vol. 45, no. 13, July 2017, pp. 7897–908. *PubMed Central*, <https://doi.org/10.1093/nar/gkx490>.
81. Koonin, Eugene V., et al. "Diversity, Classification and Evolution of CRISPR-Cas Systems." *Current Opinion in Microbiology*, vol. 37, June 2017, pp. 67–78. *PubMed Central*, <https://doi.org/10.1016/j.mib.2017.05.008>.
82. Krumova, Katerina, and Gonzalo Cosa. *Overview of Reactive Oxygen Species*. *books.rsc.org*, <https://books.rsc.org/books/edited-volume/1491/chapter/996688/Overview-of-Reactive-Oxygen-Species>. Accessed 27 July 2025.
83. Kumawat, B. L., et al. "Wilson's Disease Presenting as Isolated Obsessive-Compulsive Disorder." *Indian Journal of Medical Sciences*, vol. 61, no. 11, Nov. 2007, pp. 607–10.
84. Kunin, Victor, et al. "Evolutionary Conservation of Sequence and Secondary Structures in CRISPR Repeats." *Genome Biology*, vol. 8, no. 4, 2007, p. R61. *PubMed*, <https://doi.org/10.1186/gb-2007-8-4-r61>.

85. Lee, Choongho. "CRISPR/Cas9-Based Antiviral Strategy: Current Status and the Potential Challenge." *Molecules*, vol. 24, no. 7, Apr. 2019, p. 1349. *PubMed Central*, <https://doi.org/10.3390/molecules24071349>.
86. Lee, David Uihwan, et al. "The Cost Implications of Wilson Disease among Hospitalized Patients: Analysis of US Hospitals." *European Journal of Gastroenterology & Hepatology*, vol. 36, no. 7, July 2024, pp. 929–40. *PubMed Central*, <https://doi.org/10.1097/MEG.0000000000002777>.
87. Lenartowicz, Małgorzata, and Wojciech Krzeptowski. "[Structure and function of ATP7A and ATP7B proteins--Cu-transporting ATPases]." *Postepy Biochemii*, vol. 56, no. 3, 2010, pp. 317–27.
88. Li, Guanglei, et al. "Highly Efficient and Precise Base Editing in Discarded Human Triprenuclear Embryos." *Protein & Cell*, vol. 8, no. 10, Oct. 2017, pp. 776–79. *PubMed Central*, <https://doi.org/10.1007/s13238-017-0458-7>.
89. Liang, Puping, et al. "Correction of β -Thalassemia Mutant by Base Editor in Human Embryos." *Protein & Cell*, vol. 8, no. 11, Nov. 2017, pp. 811–22. *PubMed Central*, <https://doi.org/10.1007/s13238-017-0475-6>.
90. Lin, Steven, et al. "Enhanced Homology-Directed Human Genome Engineering by Controlled Timing of CRISPR/Cas9 Delivery." *eLife*, vol. 3, p. e04766. *PubMed Central*, <https://doi.org/10.7554/eLife.04766>.
91. Linder, MC, et al. "Copper Transport." *The American Journal of Clinical Nutrition*, vol. 67, no. 5, May 1998, pp. 965S–971S. *ScienceDirect*, <https://doi.org/10.1093/ajcn/67.5.965S>.
92. Lino, Christopher A., et al. "Delivering CRISPR: A Review of the Challenges and Approaches." *Drug Delivery*, vol. 25, no. 1, May 2018, pp. 1234–57. *PubMed Central*, <https://doi.org/10.1080/10717544.2018.1474964>.
93. Litwin, Tomasz, et al. "Psychiatric Disturbances as a First Clinical Symptom of Wilson's Disease - Case Report." *Psychiatria Polska*, vol. 50, no. 2, 2016, pp. 337–44. *PubMed*, <https://doi.org/10.12740/PP/45218>.
94. Liu, Mingjie, et al. "Methodologies for Improving HDR Efficiency." *Frontiers in Genetics*, vol. 9, Jan. 2019, p. 691. *PubMed Central*, <https://doi.org/10.3389/fgene.2018.00691>.
95. Longenecker, Joseph C., et al. "Hyperuricemia after Orthotopic Liver Transplantation: Divergent Associations with Progression of Renal Disease, Incident End-Stage Renal Disease, and Mortality." *BMC Nephrology*, vol. 18, no. 1, Mar. 2017, p. 103. *PubMed*, <https://doi.org/10.1186/s12882-017-0518-5>.
96. Lorincz, Matthew T. "Neurologic Wilson's Disease." *Annals of the New York Academy of Sciences*, vol. 1184, Jan. 2010, pp. 173–87. *PubMed*, <https://doi.org/10.1111/j.1749-6632.2009.05109.x>.
97. Lu, Yuming, and Jian-Kang Zhu. "Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System." *Molecular Plant*, vol. 10, no. 3, Mar. 2017, pp.

- 523–25. *PubMed*, <https://doi.org/10.1016/j.molp.2016.11.013>.
98. Lutsenko, Svetlana, et al. “Function and Regulation of Human Copper-Transporting ATPases.” *Physiological Reviews*, vol. 87, no. 3, July 2007, pp. 1011–46. *journals.physiology.org (Atypon)*, <https://doi.org/10.1152/physrev.00004.2006>.
 99. M, Asmamaw, and Zawdie B. “Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing.” *Biologics : Targets & Therapy*, vol. 15, Aug. 2021. *pubmed.ncbi.nlm.nih.gov*, <https://doi.org/10.2147/BTT.S326422>.
 100. M Fetherolf, Morgan, et al. “Oxygen-Dependent Activation of Cu,Zn-Superoxide Dismutase-1.” *Metallomics: Integrated Biometal Science*, vol. 9, no. 8, Aug. 2017, pp. 1047–59. *PubMed*, <https://doi.org/10.1039/c6mt00298f>.
 101. Ma, Hanhui, et al. “CRISPR-Cas9 Nuclear Dynamics and Target Recognition in Living Cells.” *Journal of Cell Biology*, vol. 214, no. 5, Aug. 2016, pp. 529–37. *Silverchair*, <https://doi.org/10.1083/jcb.201604115>.
 102. Makarova, Kira S., L. Aravind, et al. “A DNA Repair System Specific for Thermophilic Archaea and Bacteria Predicted by Genomic Context Analysis.” *Nucleic Acids Research*, vol. 30, no. 2, Jan. 2002, pp. 482–96. *PubMed*, <https://doi.org/10.1093/nar/30.2.482>.
 103. Makarova, Kira S., Nick V. Grishin, et al. “A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action.” *Biology Direct*, vol. 1, Mar. 2006, p. 7. *PubMed*, <https://doi.org/10.1186/1745-6150-1-7>.
 104. Makarova, Kira S, et al. “Unification of Cas Protein Families and a Simple Scenario for the Origin and Evolution of CRISPR-Cas Systems.” *Biology Direct*, vol. 6, July 2011, p. 38. *PubMed Central*, <https://doi.org/10.1186/1745-6150-6-38>.
 105. Marraffini, Luciano A., and Erik J. Sontheimer. “Self versus Non-Self Discrimination during CRISPR RNA-Directed Immunity.” *Nature*, vol. 463, no. 7280, Jan. 2010, pp. 568–71. *www.nature.com*, <https://doi.org/10.1038/nature08703>.
 106. Maude, Shannon L., et al. “Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia.” *The New England Journal of Medicine*, vol. 371, no. 16, Oct. 2014, pp. 1507–17. *PubMed Central*, <https://doi.org/10.1056/NEJMoa1407222>.
 107. McCord, J. M., and I. Fridovich. “Superoxide Dismutase. An Enzymic Function for Erythrocyte (Hemocuprein).” *The Journal of Biological Chemistry*, vol. 244, no. 22, Nov. 1969, pp. 6049–55.
 108. Medici, Valentina, and Janine M. LaSalle. “Genetics and Epigenetic Factors of Wilson Disease.” *Annals of Translational Medicine*, vol. 7, no. Suppl 2, Apr. 2019, p. S58. *PubMed Central*, <https://doi.org/10.21037/atm.2019.01.67>.
 109. Menerey, K. A., et al. “The Arthropathy of Wilson’s Disease: Clinical and Pathologic Features.” *The Journal of Rheumatology*, vol. 15, no. 2, Feb. 1988, pp. 331–37.
 110. Mojica, F. J., C. Díez-Villaseñor, et al. “Biological Significance of a Family of

- Regularly Spaced Repeats in the Genomes of Archaea, Bacteria and Mitochondria.” *Molecular Microbiology*, vol. 36, no. 1, Apr. 2000, pp. 244–46. *PubMed*, <https://doi.org/10.1046/j.1365-2958.2000.01838.x>.
111. Mojica, F. J., G. Juez, et al. “Transcription at Different Salinities of *Haloferax* Mediterranei Sequences Adjacent to Partially Modified PstI Sites.” *Molecular Microbiology*, vol. 9, no. 3, Aug. 1993, pp. 613–21. *PubMed*, <https://doi.org/10.1111/j.1365-2958.1993.tb01721.x>.
 112. Mojica, Francisco J. M., et al. “Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements.” *Journal of Molecular Evolution*, vol. 60, no. 2, Feb. 2005, pp. 174–82. *PubMed*, <https://doi.org/10.1007/s00239-004-0046-3>.
 113. Moreno, J. M., et al. “Chronic Renal Dysfunction after Liver Transplantation in Adult Patients: Prevalence, Risk Factors, and Impact on Mortality.” *Transplantation Proceedings*, vol. 35, no. 5, Aug. 2003, pp. 1907–08. *PubMed*, [https://doi.org/10.1016/s0041-1345\(03\)00642-0](https://doi.org/10.1016/s0041-1345(03)00642-0).
 114. Musunuru, Kiran, et al. “Patient-Specific In Vivo Gene Editing to Treat a Rare Genetic Disease.” *New England Journal of Medicine*, vol. 392, no. 22, June 2025, pp. 2235–43. *Taylor and Francis+NEJM*, <https://doi.org/10.1056/NEJMoa2504747>.
 115. Musunuru. “The Hope and Hype of CRISPR-Cas9 Genome Editing: A Review.” *JAMA Cardiology*, vol. 2, no. 8, Aug. 2017, pp. 914–19. *PubMed*, <https://doi.org/10.1001/jamacardio.2017.1713>.
 116. Neal, D. A., et al. “Hyperuricemia, Gout, and Renal Function after Liver Transplantation.” *Transplantation*, vol. 72, no. 10, Nov. 2001, pp. 1689–91. *PubMed*, <https://doi.org/10.1097/00007890-200111270-00021>.
 117. Nishida, Keiji, et al. “Targeted Nucleotide Editing Using Hybrid Prokaryotic and Vertebrate Adaptive Immune Systems.” *Science (New York, N.Y.)*, vol. 353, no. 6305, Sept. 2016, p. aaf8729. *PubMed*, <https://doi.org/10.1126/science.aaf8729>.
 118. Ogun, Aminat S., and Adebayo Adeyinka. “Biochemistry, Transferrin.” *StatPearls*, StatPearls Publishing, 2025. *PubMed*, <http://www.ncbi.nlm.nih.gov/books/NBK532928/>.
 119. Petrukhin, K., et al. “Mapping, Cloning and Genetic Characterization of the Region Containing the Wilson Disease Gene.” *Nature Genetics*, vol. 5, no. 4, Dec. 1993, pp. 338–43. *PubMed*, <https://doi.org/10.1038/ng1293-338>.
 120. Pfeifferberger, Jan, et al. “Hepatobiliary Malignancies in Wilson Disease.” *Liver International: Official Journal of the International Association for the Study of the Liver*, vol. 35, no. 5, May 2015, pp. 1615–22. *PubMed*, <https://doi.org/10.1111/liv.12727>.
 121. Pfeiffer, Ronald F. “Wilson’s Disease.” *Seminars in Neurology*, vol. 27, no. 2, Apr. 2007, pp. 123–32. *PubMed*, <https://doi.org/10.1055/s-2007-971173>.
 122. Portala, K., et al. “Psychopathology in Treated Wilson’s Disease Determined by Means of CPRS Expert and Self-Ratings.” *Acta Psychiatrica Scandinavica*, vol. 101, no. 2, Feb. 2000, pp. 104–09. *PubMed*, <https://doi.org/10.1034/j.1600-0447.2000.90085.x>.

123. Posadas, Yanahi, et al. "Copper Binding and Protein Aggregation: A Journey from the Brain to the Human Lens." *RSC Chemical Biology*, vol. 4, no. 12, pp. 974–85. *PubMed Central*, <https://doi.org/10.1039/d3cb00145h>.
124. Pourcel, C., et al. "CRISPR Elements in *Yersinia Pestis* Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies." *Microbiology (Reading, England)*, vol. 151, no. Pt 3, Mar. 2005, pp. 653–63. *PubMed*, <https://doi.org/10.1099/mic.0.27437-0>.
125. Prasad, Durga, and Arpita Bhriguvanshi. "Ocular Manifestations of Liver Disease in Children: Clinical Aspects and Implications." *Annals of Hepatology*, vol. 19, no. 6, 2020, pp. 608–13. *PubMed*, <https://doi.org/10.1016/j.aohep.2019.11.009>.
126. Pruthi, J., et al. "Analysis of Causes of Death in Liver Transplant Recipients Who Survived More than 3 Years." *Liver Transplantation: Official Publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society*, vol. 7, no. 9, Sept. 2001, pp. 811–15. *PubMed*, <https://doi.org/10.1053/jlts.2001.27084>.
127. Pugh, R N. "Pugh's Grading in the Classification of Liver Decompensation." *Gut*, vol. 33, no. 11, Nov. 1992, p. 1583. *PubMed Central*, <https://doi.org/10.1136/gut.33.11.1583>.
128. Rajalakshmi, T. R., et al. "DNA Adducts-Chemical Addons." *Journal of Pharmacy & Bioallied Sciences*, vol. 7, no. Suppl 1, Apr. 2015, pp. S197-199. *PubMed*, <https://doi.org/10.4103/0975-7406.155901>.
129. Ralle, Martina, et al. "Wilson Disease at a Single Cell Level: Intracellular Copper Trafficking Activates Compartment-Specific Responses in Hepatocytes." *The Journal of Biological Chemistry*, vol. 285, no. 40, Oct. 2010, pp. 30875–83. *PubMed*, <https://doi.org/10.1074/jbc.M110.114447>.
130. Rees, Holly A., et al. "Improving the DNA Specificity and Applicability of Base Editing through Protein Engineering and Protein Delivery." *Nature Communications*, vol. 8, June 2017, p. 15790. *pmc.ncbi.nlm.nih.gov*, <https://doi.org/10.1038/ncomms15790>.
131. Roberts, Eve A., et al. "Diagnosis and Treatment of Wilson Disease: An Update." *Hepatology (Baltimore, Md.)*, vol. 47, no. 6, June 2008, pp. 2089–111. *PubMed*, <https://doi.org/10.1002/hep.22261>.
132. Rodriguez-Granillo, Agustina, et al. "Stability and ATP Binding of the Nucleotide-Binding Domain of the Wilson Disease Protein: Effect of the Common H1069Q Mutation." *Journal of Molecular Biology*, vol. 383, no. 5, Nov. 2008, pp. 1097–111. *PubMed*, <https://doi.org/10.1016/j.jmb.2008.07.065>.
133. Román-Rodríguez, Francisco José, et al. "NHEJ-Mediated Repair of CRISPR-Cas9-Induced DNA Breaks Efficiently Corrects Mutations in HSPCs from Patients with Fanconi Anemia." *Cell Stem Cell*, vol. 25, no. 5, Nov. 2019, pp. 607-621.e7. *PubMed*, <https://doi.org/10.1016/j.stem.2019.08.016>.
134. Ross, M. E., et al. "Late-Onset Wilson's Disease with Neurological Involvement in

- the Absence of Kayser-Fleischer Rings.” *Annals of Neurology*, vol. 17, no. 4, Apr. 1985, pp. 411–13. *PubMed*, <https://doi.org/10.1002/ana.410170421>.
135. Rustgi, Vinod K., et al. “Wilson’s Disease: An Analysis of Health Care Use and Cost Burden of Commercially Insured Adults in the United States.” *Hepatology Communications*, vol. 6, no. 2, Feb. 2022, pp. 389–98. *PubMed*, <https://doi.org/10.1002/hep4.1812>.
 136. Sandahl, Thomas Damgaard, et al. “The Prevalence of Wilson’s Disease: An Update.” *Hepatology (Baltimore, Md.)*, vol. 71, no. 2, Feb. 2020, pp. 722–32. *PubMed*, <https://doi.org/10.1002/hep.30911>.
 137. Sapranaukas, Rimantas, et al. “The Streptococcus Thermophilus CRISPR/Cas System Provides Immunity in Escherichia Coli.” *Nucleic Acids Research*, vol. 39, no. 21, Nov. 2011, pp. 9275–82. *PubMed Central*, <https://doi.org/10.1093/nar/gkr606>.
 138. Sazinsky, Matthew H., et al. “Structure of the ATP Binding Domain from the Archaeoglobus Fulgidus Cu⁺-ATPase.” *The Journal of Biological Chemistry*, vol. 281, no. 16, Apr. 2006, pp. 11161–66. *PubMed*, <https://doi.org/10.1074/jbc.M510708200>.
 139. Scheiber, Ivo Florin, et al. “Pathogenesis of Wilson Disease.” *Handbook of Clinical Neurology*, vol. 142, 2017, pp. 43–55. *PubMed*, <https://doi.org/10.1016/B978-0-444-63625-6.00005-7>.
 140. Schilsky, Michael L., Eve A. Roberts, et al. “A Multidisciplinary Approach to the Diagnosis and Management of Wilson Disease: Executive Summary of the 2022 Practice Guidance on Wilson Disease from the American Association for the Study of Liver Diseases.” *Hepatology (Baltimore, Md.)*, vol. 77, no. 4, Apr. 2023, pp. 1428–55. *PubMed*, <https://doi.org/10.1002/hep.32805>.
 141. Schilsky, Michael L., Anna Czlonkowska, et al. “Trientine Tetrahydrochloride versus Penicillamine for Maintenance Therapy in Wilson Disease (CHELATE): A Randomised, Open-Label, Non-Inferiority, Phase 3 Trial.” *The Lancet. Gastroenterology & Hepatology*, vol. 7, no. 12, Dec. 2022, pp. 1092–102. *PubMed*, [https://doi.org/10.1016/S2468-1253\(22\)00270-9](https://doi.org/10.1016/S2468-1253(22)00270-9).
 142. Seniów, Joanna, et al. “Self-Rated Emotional Functioning of Patients With Neurological or Asymptomatic Form of Wilson’s Disease.” *The Clinical Neuropsychologist*, vol. 17, no. 3, Aug. 2003, pp. 367–73. *Taylor and Francis+NEJM*, <https://doi.org/10.1076/clin.17.3.367.18085>.
 143. Shalem, Ophir, et al. “Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells.” *Science (New York, N.Y.)*, vol. 343, no. 6166, Jan. 2014, pp. 84–87. *PubMed Central*, <https://doi.org/10.1126/science.1247005>.
 144. Shi, K., et al. “Structural Basis for Targeted DNA Cytosine Deamination and Mutagenesis by APOBEC3A and APOBEC3B.” *Nature Structural & Molecular Biology*, vol. 24, no. 2, Feb. 2017, pp. 131–39. *PubMed Central*, <https://doi.org/10.1038/nsmb.3344>.
 145. Shmakov, Sergey, et al. “Diversity and Evolution of Class 2 CRISPR–Cas Systems.”

- Nature Reviews. Microbiology*, vol. 15, no. 3, Mar. 2017, pp. 169–82. *PubMed Central*, <https://doi.org/10.1038/nrmicro.2016.184>.
146. Srinivas, K., et al. “Dominant Psychiatric Manifestations in Wilson’s Disease: A Diagnostic and Therapeutic Challenge!” *Journal of the Neurological Sciences*, vol. 266, nos. 1–2, Mar. 2008, pp. 104–08. *PubMed*, <https://doi.org/10.1016/j.jns.2007.09.009>.
 147. Sternberg, Samuel H., et al. “DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9.” *Nature*, vol. 507, no. 7490, Mar. 2014, pp. 62–67. *PubMed*, <https://doi.org/10.1038/nature13011>.
 148. Stohs: *Oxidative Mechanisms in the Toxicity of Metal Ions* - *Google Scholar*. https://scholar.google.com/scholar_lookup?title=Oxidative%20mechanisms%20in%20the%20toxicity%20of%20metal%20ions&publication_year=1995&author=S.J.%20Stohs&author=D.%20Bagchi. Accessed 27 July 2025.
 149. Sturniolo, Giacomo Carlo, et al. “Small Fiber Peripheral Neuropathy in Wilson Disease: An In Vivo Documentation by Corneal Confocal Microscopy.” *Investigative Ophthalmology & Visual Science*, vol. 56, no. 2, Feb. 2015, pp. 1390–95. *Silverchair*, <https://doi.org/10.1167/iovs.14-15004>.
 150. Sussman, W., and I. H. Scheinberg. “Disappearance of Kayser-Fleischer Rings. Effects of Penicillamine.” *Archives of Ophthalmology (Chicago, Ill.: 1960)*, vol. 82, no. 6, Dec. 1969, pp. 738–41. *PubMed*, <https://doi.org/10.1001/archopht.1969.00990020730003>.
 151. Svetel, M., et al. “Long-Term Outcome in Serbian Patients with Wilson Disease.” *European Journal of Neurology*, vol. 16, no. 7, July 2009, pp. 852–57. *PubMed*, <https://doi.org/10.1111/j.1468-1331.2009.02607.x>.
 152. Svetel, Marina, et al. “Neuropsychiatric Aspects of Treated Wilson’s Disease.” *Parkinsonism & Related Disorders*, vol. 15, no. 10, Dec. 2009, pp. 772–75. *ScienceDirect*, <https://doi.org/10.1016/j.parkreldis.2009.01.010>.
 153. Szczelkun, Mark D., et al. “Direct Observation of R-Loop Formation by Single RNA-Guided Cas9 and Cascade Effector Complexes.” *Proceedings of the National Academy of Sciences*, vol. 111, no. 27, July 2014, pp. 9798–803. *pnas.org (Atypon)*, <https://doi.org/10.1073/pnas.1402597111>.
 154. Tang, Weixin, et al. “Aptazyme-Embedded Guide RNAs Enable Ligand-Responsive Genome Editing and Transcriptional Activation.” *Nature Communications*, vol. 8, June 2017, p. 15939. *PubMed Central*, <https://doi.org/10.1038/ncomms15939>.
 155. Tanzi, R. E., et al. “The Wilson Disease Gene Is a Copper Transporting ATPase with Homology to the Menkes Disease Gene.” *Nature Genetics*, vol. 5, no. 4, Dec. 1993, pp. 344–50. *PubMed*, <https://doi.org/10.1038/ng1293-344>.
 156. Tapiero, H., et al. “Trace Elements in Human Physiology and Pathology. Copper.” *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, vol. 57, no. 9, Nov. 2003, pp. 386–98. *PubMed Central*, [https://doi.org/10.1016/s0753-3322\(03\)00012-x](https://doi.org/10.1016/s0753-3322(03)00012-x).

157. Terada, K., et al. "ATP7B (WND) Protein." *The International Journal of Biochemistry & Cell Biology*, vol. 30, no. 10, Oct. 1998, pp. 1063–67. *PubMed*, [https://doi.org/10.1016/s1357-2725\(98\)00073-9](https://doi.org/10.1016/s1357-2725(98)00073-9).
158. Terns, Michael P., and Rebecca M. Terns. "CRISPR-Based Adaptive Immune Systems." *Current Opinion in Microbiology*, vol. 14, no. 3, June 2011, pp. 321–27. *PubMed Central*, <https://doi.org/10.1016/j.mib.2011.03.005>.
159. Verity, M. Anthony, and J. K. Gambell. "Studies of Copper Ion-Induced Mitochondrial Swelling in Vitro." *Biochemical Journal*, vol. 108, no. 2, June 1968, pp. 289–95. *Silverchair*, <https://doi.org/10.1042/bj1080289>.
160. Walshe, J. M. "The Acute Haemolytic Syndrome in Wilson's Disease--a Review of 22 Patients." *QJM: Monthly Journal of the Association of Physicians*, vol. 106, no. 11, Nov. 2013, pp. 1003–08. *PubMed*, <https://doi.org/10.1093/qjmed/hct137>.
161. Walshe, J. M., and G. Potter. "The Pattern of the Whole Body Distribution of Radioactive Copper (67Cu, 64Cu) in Wilson's Disease and Various Control Groups." *The Quarterly Journal of Medicine*, vol. 46, no. 184, Oct. 1977, pp. 445–62.
162. Wang, Jianbin, and Stephen R. Quake. "RNA-Guided Endonuclease Provides a Therapeutic Strategy to Cure Latent Herpesviridae Infection." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 36, Sept. 2014, pp. 13157–62. *PubMed Central*, <https://doi.org/10.1073/pnas.1410785111>.
163. Wang, Karissa, and Matthew Nicholaou. "Suppression of Antimicrobial Resistance in MRSA Using CRISPR-dCas9." *American Society for Clinical Laboratory Science*, vol. 30, no. 4, Oct. 2017, pp. 207–13. Research and Reports. [clsjournal.ascls.org](https://doi.org/10.29074/ascls.30.4.207), <https://doi.org/10.29074/ascls.30.4.207>.
164. Wang, Ming, et al. "Efficient Delivery of Genome-Editing Proteins Using Bio-reducible Lipid Nanoparticles." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 11, Mar. 2016, pp. 2868–73. *PubMed Central*, <https://doi.org/10.1073/pnas.1520244113>.
165. Weiss, Karl Heinz, Mart Van de Moortele, et al. "Bone Demineralisation in a Large Cohort of Wilson Disease Patients." *Journal of Inherited Metabolic Disease*, vol. 38, no. 5, Sept. 2015, pp. 949–56. *PubMed*, <https://doi.org/10.1007/s10545-015-9815-y>.
166. Weiss, Karl Heinz, Florentine Thurik, et al. "Efficacy and Safety of Oral Chelators in Treatment of Patients with Wilson Disease." *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, vol. 11, no. 8, Aug. 2013, pp. 1028-1035.e1-2. *PubMed*, <https://doi.org/10.1016/j.cgh.2013.03.012>.
167. Weiss, Karl Heinz, Mark Schäfer, et al. "Outcome and Development of Symptoms after Orthotopic Liver Transplantation for Wilson Disease." *Clinical Transplantation*, vol. 27, no. 6, 2013, pp. 914–22. *PubMed*, <https://doi.org/10.1111/ctr.12259>.
168. Weiss, Karl Heinz, and Michael Schilsky. "Wilson Disease." *GeneReviews®*, edited by Margaret P. Adam et al., University of Washington, Seattle, 1993. *PubMed*,

- <http://www.ncbi.nlm.nih.gov/books/NBK1512/>.
169. Wiebers, D. O., et al. "The Ophthalmologic Manifestations of Wilson's Disease." *Mayo Clinic Proceedings*, vol. 52, no. 7, July 1977, pp. 409–16. *PubMed*, <https://pubmed.ncbi.nlm.nih.gov/875462/>.
 170. Wiedenheft, Blake, et al. "RNA-Guided Genetic Silencing Systems in Bacteria and Archaea." *Nature*, vol. 482, no. 7385, Feb. 2012, pp. 331–38. *PubMed*, <https://doi.org/10.1038/nature10886>.
 171. Wong, P. C., et al. "Copper Chaperone for Superoxide Dismutase Is Essential to Activate Mammalian Cu/Zn Superoxide Dismutase." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 6, Mar. 2000, pp. 2886–91. *PubMed*, <https://doi.org/10.1073/pnas.040461197>.
 172. Yang, Han, et al. "Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks." *International Journal of Molecular Sciences*, vol. 21, no. 18, Sept. 2020, p. 6461. *PubMed Central*, <https://doi.org/10.3390/ijms21186461>.
 173. Yin, Hao, et al. "Therapeutic Genome Editing by Combined Viral and Non-Viral Delivery of CRISPR System Components in Vivo." *Nature Biotechnology*, vol. 34, no. 3, Mar. 2016, pp. 328–33. *PubMed Central*, <https://doi.org/10.1038/nbt.3471>.
 174. Zhen, S., et al. "Harnessing the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-Associated Cas9 System to Disrupt the Hepatitis B Virus." *Gene Therapy*, vol. 22, no. 5, May 2015, pp. 404–12. *PubMed*, <https://doi.org/10.1038/gt.2015.2>.
 175. Zhou, Yuexin, et al. "High-Throughput Screening of a CRISPR/Cas9 Library for Functional Genomics in Human Cells." *Nature*, vol. 509, no. 7501, May 2014, pp. 487–91. *PubMed*, <https://doi.org/10.1038/nature13166>.
 176. Zhuang, Xiao-Hui, et al. "Analysis of Renal Impairment in Children with Wilson's Disease." *World Journal of Pediatrics: WJP*, vol. 4, no. 2, May 2008, pp. 102–05. *PubMed*, <https://doi.org/10.1007/s12519-008-0019-5>.
 177. Zimbrea, Paula C., and Michael L. Schilsky. "Psychiatric Aspects of Wilson Disease: A Review." *General Hospital Psychiatry*, vol. 36, no. 1, 2014, pp. 53–62. *PubMed*, <https://doi.org/10.1016/j.genhosppsych.2013.08.007>.
 178. Zischka, Hans, et al. "Liver Mitochondrial Membrane Crosslinking and Destruction in a Rat Model of Wilson Disease." *The Journal of Clinical Investigation*, vol. 121, no. 4, Apr. 2011, pp. 1508–18. www.jci.org, <https://doi.org/10.1172/JCI45401>.
 179. Zong, Yuan, et al. "Precise Base Editing in Rice, Wheat and Maize with a Cas9-Cytidine Deaminase Fusion." *Nature Biotechnology*, vol. 35, no. 5, May 2017, pp. 438–40. *PubMed*, <https://doi.org/10.1038/nbt.3811>.
 180. CHANG, IRENE J., and SI HOUN HAHN. "The Genetics of Wilson Disease." *Handbook of Clinical Neurology*, vol. 142, 2017, pp. 19–34. *PubMed Central*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5648646/>.

181. Chevalier, Kevin, et al. "Eye Involvement in Wilson's Disease: A Review of the Literature." *Journal of Clinical Medicine*, vol. 11, no. 9, Apr. 2022, p. 2528. *PubMed Central*, <https://doi.org/10.3390/jcm11092528>.
182. Schilsky, Michael L., et al. "Costly Choices for Treating Wilson's Disease." *Hepatology (Baltimore, Md.)*, vol. 61, no. 4, Apr. 2015, pp. 1106–08. *PubMed*, <https://doi.org/10.1002/hep.27663>.
183. Chadwick, Alexandra C., et al. "In Vivo Base Editing of PCSK9 as a Therapeutic Alternative to Genome Editing." *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 37, no. 9, Sept. 2017, pp. 1741–47. *PubMed Central*, <https://doi.org/10.1161/ATVBAHA.117.309881>.
184. Komor, Alexis C., et al. "Editing the Genome Without Double-Stranded DNA Breaks." *ACS Chemical Biology*, vol. 13, no. 2, Feb. 2018, pp. 383–88. *PubMed Central*, <https://doi.org/10.1021/acscchembio.7b00710>.
185. Bull, P. C., et al. "The Wilson Disease Gene Is a Putative Copper Transporting P-Type ATPase Similar to the Menkes Gene." *Nature Genetics*, vol. 5, no. 4, Dec. 1993, pp. 327–37. *PubMed*, <https://doi.org/10.1038/ng1293-327>.

Legal Battles for Belonging: Chinese Immigrants and the Constitution, 1886–1927

By Tianli Cheng

Introduction

Chinese immigrants were pushed to the margins of law and society. Through state and federal measures like the Chinese Exclusion Act of 1882, the nation sought to erase their presence and silence their voices in nineteenth-century America. They faced hostility not only in mines, railroads, and communities but also in the very structures of government. Their fight also extended from outside the court to inside. The courtrooms became one of the spaces where Chinese immigrants could argue their unfairness, fight for their status, and assert their dignity. Through their contest, they not only forced the nation to confront fundamental questions about citizenship, human rights, equality, but also on the broader meaning of constitutional protections for everyone in the United States.

This paper traces those legal battles from *Yick Wo v. Hopkins* (1886) to *United States v. Wong Kim Ark* (1898), and through the early twentieth century, with the cases of *Lum v. Rice* (1927). The lawsuits that Chinese immigrants brought exposed the contradictions of American democracy. Decisions such as *Chae Chan Ping* and *Fong Yue Ting* reinforced exclusion and gave Congress sweeping deportation powers, while *Yick Wo* and *Wong Kim Ark* affirmed principles of equal protection and birthright citizenship that still shape constitutional law today. By following this arc, the essay argues that Chinese immigrants, though denied political voice, reshaped the constitutional meaning of race, citizenship, and belonging in the United States.

Equal Protection in Practice: *Yick Wo v. Hopkins* (1886) and its Impact

Chinese immigrants who came to the United States in the mid-19th century in pursuit of their American dreams found themselves confined not only by hostility in the streets but also by a steady stream of local ordinances designed to push them out of economic life. One of the most notorious ones is the regulation targeting laundries issued by the city of San Francisco, a business that many Chinese had found modest success in. In 1880, San Francisco passed an ordinance requiring laundry owners operating in wooden buildings to obtain a permit from the Board of Supervisors. On paper, the law appeared neutral, but in practice, its enforcement was starkly discriminatory. Out of roughly two hundred Chinese applicants, not a single one was granted a permit, while nearly all white applicants received approval (McClain 102–105).

Among those denied was Yick Wo, the operator of a wooden-frame laundry who had complied with all safety requirements. Refusing to close his business, he was fined and jailed. Instead of accepting the punishment, he challenged the ordinance in court. His case, joined by another Chinese laundryman, Wo Lee, eventually reached the United States Supreme Court. Arguing before the justices, Yick Wo's lawyers argued that the ordinance was a sham of equality. A law that looked neutral in wording on the page, they explained, in practice, it stripped Chinese laundrymen of rights the Fourteenth Amendment promised to *every person* living under U.S. law, regardless of citizenship (U.S. Const. amend. XIV, sec. 1). The city of San Francisco

countered that municipalities had broad discretion to regulate businesses for public safety and that courts should not interfere with administrative judgments.

In a unanimous decision, the Court sided with Yick Wo. Justice Stanley Matthews declared that laws “fair on their face, and impartial in appearance” could still violate the Constitution if enforced “with an evil eye and an unequal hand.” (*Yick Wo v. Hopkins* 373). What made the decision groundbreaking was that it recognized that discrimination could be embedded not just in explicit statutes but also in the way neutral laws were applied.

The significance of *Yick Wo v. Hopkins* extended well beyond Chinese laundries. By the Court declared that the Equal Protection Clause extended to “all persons within the territorial jurisdiction, without regard to differences of race, of color, or of nationality,” it made clear that even non-citizens were entitled to constitutional safeguards (*Yick Wo v. Hopkins* 369). It is a remarkable stance in an era when Chinese immigrants were legally excluded from naturalization. This recognition was extraordinary.

However, the victory still carried limits. Although the decision established a powerful precedent for equal protection, Chinese immigrants continued to face hostile legislation and discriminatory enforcement in other domains. The ruling did not invalidate the broader system of exclusion that Congress had already enacted. Nevertheless, *Yick Wo* became one of the first Supreme Court cases. It articulated the principle that discriminatory administration of law is as unconstitutional as discriminatory law itself. Historians and legal scholars often consider the case of *Yick Wo v. Hopkins* as a foundation for later civil rights jurisprudence in the twentieth century (Chin 151–53).

The Chinese Exclusion Case: Chae Chan Ping v. United States in 1889

The optimism of *Yick Wo* (1886) did not last long. Only three years later, the Supreme Court’s ruling in *Chae Chan Ping v. United States* (1889), better known as the *Chinese Exclusion Case*, marked a sharp turn back toward exclusion, better known as the *Chinese Exclusion Case*. Chae Chan Ping had lived in San Francisco since the 1870s and, like many others, had secured a certificate under the Burlingame and Angell treaties that permitted his reentry if he traveled abroad. In 1887, when he returned from visiting China, he found himself barred by the Scott Act, which Congress had passed while he was away, and stripped validity from the very certificates that Chinese residents had been promised, shutting the door on their return (Lee 37–39). In response, Chae tried to overturn the refusal of his reentry. He argued that the Scott Act broke treaty promises and stripped him of liberty without due process. His attorneys further warned that Congress could not erase international agreements at will (McClain 187–89). However, the government pressed the Court to recognize Congress’s “plenary power” over immigration, and affirm the congressional “plenary power,” portraying exclusion as an act of national defense and security.

In siding with Congress, the Court made a unanimous decision. Justice Stephen Field wrote that the power to exclude foreigners was an inherent attribute of sovereignty, “absolute and unqualified” (*Chae Chan Ping v. United States*). Treaties, he explained, were political instruments

that Congress could override at will. By yielding so completely to the political branches, As the result, the Court placed immigration policy beyond the reach of judicial review.

For Chinese immigrants, the consequences devastated the heart of their community. Families found themselves divided, with some members who left the U.S. with a valid reentry certificate being denied return. The promises that were once guaranteed by international agreements immediately fell with the ruling, such as the Burlingame Treaty of 1868, which established friendly relations between the U.S. and China, affirming the right of free migration. Also, the Angell Treaty of 1880, which modified the Burlingame Treaty, gave the U.S. the power to regulate, limit, or suspend immigration of Chinese laborers, but still protected the rights of those already in the country. Both treaties collapsed overnight with the passage of the Scott Act. Historians estimated that the number exceeded 20,000 who had left with lawful reentry papers were permanently barred from returning (Salyer 79). The decision cemented a constitutional doctrine that would haunt immigrants long after: that Congress's authority over admission and exclusion was virtually unchecked.

The decision in *Chae Chan Ping* could not have been more different from *Yick Wo*. In San Francisco, Chinese immigrants had persuaded the Court that equal protection meant something even for noncitizens. At the nation's gates, however, the justices took a different stance, treating exclusion as a question of sovereignty rather than of rights. The lesson was unmistakable: within American borders, they might sometimes find recourse in law, but at the border itself, their claims were silenced (Chin 21–23).

Deportation and Sovereignty: Fong Yue Ting v. United States (1893)

If *Chae Chan Ping's* case had established the government's power to exclude newcomers, the case of *Fong Yue Ting v. United States* expanded that power inward, giving Congress nearly unchecked authority to deport those already settled. The case arose from the Geary Act of 1892, which reinforced the Chinese Exclusion Act of 1882. The Geary Act demanded that every Chinese resident needed to obtain a certificate of residence, which was endorsed by the testimony of a “credible white witness.” Failure to carry such a certificate meant arrest and deportation (Salyer 102–03).

Even though they lived in the country for years, Fong Yue Ting and two other Chinese laborers in New York had been ordered deported because they lacked the required residence certificates. Their lawyers argued that expelling residents without a jury trial violated the Fifth and Sixth Amendments, which guaranteed due process and trial by jury (McClain 192–94). Deportation, they contended, was essentially a punishment and should be treated as such under the Constitution. The government insisted otherwise. Solicitor General Aldrich told the Court that deportation was not a punishment but an administrative measure, an extension of the sovereign power to regulate the presence of foreigners (Neuman 1843). If Congress could decide who may enter, it could also decide who must leave. The Supreme Court agreed with the government in a 6–3 decision.

Justice Horace Gray, a longtime Massachusetts jurist appointed to the Court in 1881, wrote the majority opinion. He had already shaped the law of Chinese immigration in *Fong Yue Ting* (1893), where he upheld Congress's sweeping deportation powers, making his later defense of birthright citizenship all the more striking. Only five years later, He would author the opinion in *Wong Kim Ark*, writing for the majority that deportation was a civil, not criminal, proceeding. "The right of a nation to expel or deport foreigners who have not been naturalized, or taken any steps towards becoming citizens," he said, "rests upon the same grounds, and is as absolute and unqualified, as the right to prohibit and prevent their entrance into the country" (*Fong Yue Ting v. United States* 707). His words reflected that deportation was necessary, holding that Congress had plenary authority to set conditions for removal.

The consequences for Chinese immigrants were devastating. Without papers, even long-term residents who had lived peacefully in the United States could also be rounded up and expelled. In terms of the requirement "credible white witness" to obtain the certificate, it exposed the racial biases embedded in the U.S. law. As Historian Lucy Salyer pointed out, this case shows how the ruling entrenched deportation as a routine instrument of immigration control, shifting power decisively away from the courts and into the hands of administrative officers (Salyer 105–07).

In the broader arc of Chinese exclusion, *Fong Yue Ting* was a grim confirmation: Chinese immigrants were present at the sufferance of Congress, and even lawful residence gave them no security. The contrast with *Wong Kim Ark* a few years later was striking. The Court could affirm birthright citizenship for the children of Chinese immigrants, but their parents could be deported at Congress's command. In this way, citizenship itself became the dividing line between who can be protected by the Constitution and who cannot.

Citizenship and Identity: United States v. Wong Kim Ark (1898)

By the 1890s, the Chinese Exclusion Act, reinforced by a series of amendments, left little room for hope for Chinese immigrants, but the question of the status of their U.S.-born children had never been settled, and therefore forced the courts to draw new boundaries. Wong Kim Ark, one of them, was born in San Francisco in 1873, the son of parents who remained Chinese nationals, and believed his citizenship was secure by the birthplace of California. In 1895, after a temporary visit to China, he was denied reentry by the U.S. customs, claiming that the Exclusion laws stripped him of citizenship (McClain 178–80).

To respond to the government's action, Wong argued that being born in San Francisco placed him within the clear promise of the Fourteenth Amendment. His attorneys grounded their argument on the Fourteenth Amendment and directly quoted the exact words, "All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States" (U.S. Const. amend. XIV, sec.1) The government, by contrast, argued that the amendment did not extend to children of Chinese immigrants, since their parents owed allegiance to a foreign sovereign and were barred from naturalization.

In 1898, the justices delivered a 6–2 decision siding with Wong. Justice Horace Gray, speaking for the majority, grounded his opinion in centuries of English common law as well as American precedent. Citizenship by birth, or *jus soli*, he explained, was a principle so deeply rooted that only a constitutional amendment could override it (Salyer 112–13). In other words, the Court rejected the government’s claims that race or parental status could limit the amendment’s scope. Gray wrote, “Every child born in the United States of parents who, at the time of his birth, are subject to the jurisdiction thereof, becomes at once a citizen of the United States”(United States v. Wong Kim Ark 693).

The decision was a rare victory for the Chinese community. It is confirmed that children born in the United States would not be left stateless, even if their parents could never naturalize. Some scholars have called the ruling a lifeline that case has been remembered as a rare safeguard, a moment when the Constitution opened its protection to Chinese Americans during an era of hostility (Lee 66–67). Even so, the victory had sharp limits. The ruling left intact the harsh exclusion system at the nation’s borders but drew a line that inside the United States, birth guaranteed citizenship; outside, exclusion still reigned.

The significance of the *Wong Kim Ark* case has only grown with time. Later debates about the citizenship of children born to immigrants from the children of Japanese immigrants in the early twentieth century to modern controversies over so-called “birthright citizenship” that all trace back to this case (Ngai 38–40). What began as the plea of one man at San Francisco’s docks became a cornerstone of constitutional law.

Segregation and Citizenship: Gong Lum v. Rice (1927)

Although the Supreme Court had already recognized birthright citizenship in *Wong Kim Ark*, the promise of equal treatment remained elusive for Chinese immigrants. By the 1920s, Chinese Americans had lived in the United States for generations. The experience of Martha Lum, a nine-year-old girl in Mississippi, was told she could not attend her local school, which made that painfully clear.

Martha was born in the United States to Chinese immigrant parents. In 1924, she tried to attend a local public school, only to be told she could not enter because she was “colored.” The school board directed her to the school designated for Black children. Her father, Gong Lum, sued, arguing that his daughter’s citizenship gave her the right to attend the white school in her district.¹

In 1927, the case reached the Supreme Court. Chief Justice William Howard Taft, writing for a unanimous Court, sided with the school board.² He leaned on *Plessy v. Ferguson* (1896), the ruling that had entrenched segregation under the banner of “separate but equal.” If Black and white children could be separated, Taft reasoned, so could Chinese children. Citizenship, in other words, offered no escape from the state’s power to enforce racial lines in education.

The Court’s ruling placed Chinese American children into the same segregated system imposed on African Americans. As Historian Charles McClain pointed out, the case is “a stark reminder that citizenship and equality were far from synonymous.”³ For the Lum family, the

decision was a bitter defeat, and for the Chinese American community, it revealed how easily the color line could override legal status.

Although *Gong Lum* is less well known than *Yick Wo* or *Wong Kim Ark*, its significance lies in its reminder that constitutional victories were fragile. Chinese Americans might secure recognition as citizens in one era, only to find themselves swept into America's entrenched system of racial segregation in another.

Short Epilogue: The Repeal of Exclusion

While the case of *Gong Lum* (1927) ends this article's case studies, the story of Chinese exclusion did not end there. During World War II, when China became an ally with the U.S. in the fight against Japan, Congress finally repealed the Exclusion Laws through the Magnuson Act of 1943, which set a token quota of 105 Chinese immigrants a year. The measure was limited, but it signaled a shift. It showed how wartime alliances and international pressure forced Americans to reconsider old policies, even if discrimination lingered (Lee 280–82).

Until the landmark Immigration and Nationality Act of 1965, in the decades that followed, Chinese Americans, like other minority groups, also benefited indirectly from the broader civil rights movement during the 1950s and 1960s. However, the century-long legal battles waged by Chinese immigrants from *People v. Hall* to the Magnuson Act left an enduring legacy that was a constitutional foundation that later generations would build upon in their pursuit of equality and belonging.

Conclusion

From the California goldfields of the 1850s to Martha Lum's exclusion from a white school in the 1920s, Chinese immigrants and their children encountered a legal system that seemed determined to keep them on the margins. *People v. Hall* silenced their testimony; *Chae Chan Ping* and *Fong Yue Ting* confirmed Congress's sweeping power over immigration. Yet other rulings, such as *Yick Wo* and *Wong Kim Ark*, affirmed principles of equal protection and birthright citizenship that remain central to American law.

The contradictions were sharp. The same Court that secured citizenship for children born in the United States also upheld the deportation of their parents. These cases remind us that rights in America have rarely been granted all at once; they were uneven, contested, and often defined by race.

What stands out is the persistence of Chinese immigrants who sought justice in the courts despite overwhelming odds. Through the process, they forced the nation to directly confront the hard questions about what equality, rights, and belonging should be. By fighting, Chinese immigrants reshaped the constitutional law, leaving precedents that continue to guide debates over immigration and citizenship today.

Works Cited

- Chae Chan Ping v. United States, 130 U.S. 581 (1889). Justia U.S. Supreme Court Center, <https://supreme.justia.com/cases/federal/us/130/581/>.
- Chin, Gabriel J. "The Plenary Power Doctrine." Immigration and Nationality Law Handbook, edited by Austin T. Fragomen Jr., American Immigration Lawyers Association, 2005, pp. 21–23.
- Chin, Gabriel J. "The Plessy Myth: Justice Harlan and the Chinese Cases." Iowa Law Review, vol. 82, no. 1, 1996, pp. 151–203.
- Fong Yue Ting v. United States, 149 U.S. 698 (1893).
- Lee, Erika. At America's Gates: Chinese Immigration during the Exclusion Era, 1882–1943. University of North Carolina Press, 2003.
<https://uncpress.org/9780807854488/at-americas-gates/>.
- McClain, Charles J. In Search of Equality: The Chinese Struggle Against Discrimination in Nineteenth-Century America. University of California Press, 1994.
- Ngai, Mae M. Impossible Subjects: Illegal Aliens and the Making of Modern America. Princeton University Press, 2004.
- Neuman, Gerald L. "The Lost Century of American Immigration Law (1776–1875)." Columbia Law Review, vol. 93, no. 8, 1993, pp. 1833–1901.
- Salyer, Lucy E. Laws Harsh as Tigers: Chinese Immigrants and the Shaping of Modern Immigration Law. University of North Carolina Press, 1995.
- U.S. Constitution. Amendment XIV, Section 1. Legal Information Institute, Cornell Law School, <https://www.law.cornell.edu/constitution/amendmentxiv>.
- United States v. Wong Kim Ark, 169 U.S. 649 (1898).
- Yick Wo v. Hopkins, 118 U.S. 356 (1886). Legal Information Institute, Cornell Law School, <https://www.law.cornell.edu/supremecourt/text/118/356>.

Navigating Economic Recovery Post-COVID-19: The Role of Policy, Labor Flexibility, Healthcare, and Digital Infrastructure By Shreemay Shah

Abstract

The COVID-19 pandemic had a major impact on the global economy, leading to a significant downturn and forcing governments to respond with various economic policies. Research shows that factors like fiscal and monetary policies, labor market flexibility, the size of the healthcare sector, and access to the internet played a key role in how well countries recovered. Growth rates from 2012 to 2024 varied widely, but a study of 25 countries found that when monetary and fiscal policies were well-coordinated, they helped support economic growth. However, countries with less flexible labor markets experienced slower recoveries. Stronger healthcare systems were also linked to better economic resilience, and widespread internet access made it easier for economies to bounce back after the initial shock.

Intro and Context

The COVID-19 pandemic was one of the most disruptive events in modern history, profoundly affecting both public health and the global economy. In 2020, approximately 114 million jobs were lost worldwide, with global working hours declining by 10.5%—equivalent to 255 million full-time jobs—and resulting in \$3.7 trillion in lost labor income, according to the International Labour Organization (Richter). Business closures were widespread: nearly 1.2 million U.S. jobs were permanently eliminated due to establishment exits in the second quarter of 2020 (Decker & Haltiwanger). In response, governments enacted sweeping measures to contain the virus and stabilize economies. The United States passed the \$2 trillion CARES Act in March 2020, followed by an additional \$900 billion relief package later that year. These interventions were critical in preventing deeper economic collapse, though the long-term consequences of the pandemic continue to unfold.

Some nations recovered quickly, while others continued to struggle for years. This raised important questions: Why did some countries do better than others? What kinds of policies worked best? And what other factors—besides government actions—played a role in shaping recovery?

This report tries to answer those questions by looking at how different policies and systems affected economic growth after the pandemic. I look not only at what governments have done, but also at the underlying structure of each country's economy and society to better understand their response to the crisis. Some countries had more flexible labor markets, meaning it was easier for people to get hired or switch jobs. Others had stronger healthcare systems, which helped them manage the health crisis more effectively. And some had better internet access, which made it easier to work remotely, attend school online, and keep businesses running during lockdowns.

To study these differences, I focused on four main areas:

- **Fiscal and monetary policy:** This includes how governments spent money and how they used tools like interest rates to influence the economy. I looked at whether countries used one type of policy or both, and how that affected their growth.
- **Labor market flexibility:** This refers to how easy it is for employers to hire and fire workers, and how quickly people can find new jobs. Countries with flexible labor markets may have been able to adjust faster during the crisis.
- **Healthcare sector size:** A strong healthcare system can help a country respond to a pandemic more effectively. I looked at whether countries with larger healthcare sectors had better outcomes.
- **Internet access:** Digital tools became essential during COVID-19. Countries with better internet access were able to shift to remote work, online learning, and digital services more easily.

To understand how well countries recovered from the COVID-19 pandemic, I looked at their average economic growth rates before and after the crisis. Specifically, I compared data from 2012 to 2019 with data from 2020 to 2024, using information from the World Bank, which tracks economic performance around the world. Growth rates are a helpful way to measure recovery, as they demonstrate whether a country's economy is growing or shrinking. If a country's growth rate declined sharply after 2020, it likely faced greater challenges in recovering from the crisis. In contrast, if its growth rate remained stable or improved, that indicates the country responded to the situation more effectively.

It is important to say that I cannot prove these policies or systems directly caused the changes in growth. Every country is different, and many things affect economic performance, including population size, geography, political stability, and more. What's more, the policies or systems I study may be correlated with these distinct determinants of economic performance. However, by comparing groups of countries with similar characteristics, I can spot patterns and learn what might have helped or hurt recovery. These insights are useful for understanding the past and preparing for future challenges.

In the following sections, I will examine the four major areas impacted by the COVID-19 pandemic in greater detail. By reviewing current research and analyzing relevant data, I aim to understand the scope of these effects and evaluate how different countries responded to the crisis. This analysis will also highlight key lessons that can help guide future decision-making in times of global disruption.

Qualitative Analysis

Section 1: Fiscal and Monetary Policy

Definition and Context: Fiscal policy refers to government decisions on spending, taxation, and borrowing to influence economic activity. Monetary policy involves central bank actions such as interest rate changes, quantitative easing, and liquidity support to stabilize

inflation and promote growth. During the COVID-19 pandemic, countries deployed both aggressively to prevent economic collapse.

Importance for Recovery: Fiscal and monetary policies are extremely important during a crisis such as COVID-19. They help keep the economy stable by making sure people have money to spend, businesses can stay open, and confidence doesn't collapse (Heideri et. all).

Fiscal policy is about how the government spends money and handles taxes. During the pandemic, many countries gave direct cash payments to people—like the U.S. sending out \$1,200 and \$1,400 stimulus checks—to help cover basic needs and boost spending. Governments also used wage subsidies to help businesses keep workers employed. Canada's Emergency Wage Subsidy covers up to 75% of wages for struggling Monetary policy is controlled by central banks and focuses on interest rates and the money supply. During COVID, banks like the Federal Reserve and European Central Bank lowered interest rates to make borrowing cheaper. This helped businesses stay afloat and encouraged people to spend. Some banks also used quantitative easing, which means buying government bonds to increase money in the system. The Bank of England did this to keep markets stable and interest rates low. Additionally, Monetary policy also boosts confidence. When people see that central banks are taking action, they are more prone to spending and investing, which contributes to the economy recovering faster.

Literature Review: Studies show fiscal stimulus has strong short-term effects, especially in crises. For example, Gechert found that public investment multipliers, which measure how much economic activity is created from government spending, were 0.5 pp higher than regular spending. That means when governments invest in things like infrastructure or schools, this approach yields a more pronounced economic impact than merely distributing funds or implementing tax cuts. Hlavacek & Ismayilov looked at over 3,200 estimates and found that the average fiscal multiplier was between 0.75 and 0.82. This means that for every dollar the government spends, the economy grows by about 75 to 82 cents—or even more in some cases. That has a significant impact, especially when demand is low and people are not spending much on their own. These findings suggest that fiscal stimulus works well in a crisis, because it directly increases demand, creates jobs, and helps businesses stay open. When the private sector is struggling, government spending can fill the gap and promote economic recovery. To support this, Yuan et al. showed that in China, fiscal policy was more effective than monetary policy in post-COVID recovery. Yeboah emphasized that coordination between fiscal and monetary policy enhances macroeconomic stability. Overall, the evidence suggests that when the private sector slows down, government spending can step in to support growth and help economies recover more quickly.

Conclusion: Countries that used direct cash payments (e.g., US, Ireland) and large-scale stimulus packages saw faster GDP rebounds than those with limited fiscal space.

Section 2: Labor Market Flexibility

Definition and Context: Labor market flexibility refers to how easily employers and workers can adjust hiring, firing, wages, and work conditions in response to economic changes. Flexible markets allow businesses to adapt quickly during shocks, helping reduce unemployment and speed up recovery. In contrast, rigid markets offer more job security but can slow down economic adjustment and prolong downturns (Bernal-Verdugo et. al).

Importance for Recovery: During economic shocks, flexible labor markets help reallocate labor quickly, reduce unemployment duration, and support productivity. They also enable faster job creation during recovery phases. This adaptability allows businesses to respond to changing demand without being constrained by rigid labor regulations. As a result, economies with flexible labor systems tend to rebound more efficiently and maintain stronger long-term growth.

Literature Review: Recent research highlights how labor market structure can shape recovery outcomes. Pyati found that flexible labor markets are linked to lower unemployment and higher job creation. Although the study doesn't focus specifically on crisis periods, it suggests that flexibility helps economies reallocate labor more efficiently and respond to changing demand during recovery. Ramos et al. also showed that deregulated markets with fewer unions tend to have better wage responsiveness and stronger recovery performance, reinforcing the idea that adaptability plays a key role.

Building on this, several studies have looked at how active labor market programs (ALMPs) support employment during and after downturns. Card, Kluve, and Weber reviewed over 200 studies and found that ALMPs often have limited short-term effects but show stronger results two to three years after implementation—especially in flexible labor markets. They also noted that these programs tend to be more effective during and after recessions, making them a valuable tool for post-crisis recovery. Vooren et al. found that job-search assistance and training programs had lasting positive impacts on employment for up to 36 months. While not focused on crisis periods, these findings are especially relevant after COVID, when many workers needed support to re-enter a shifting labor market.

Conclusion: The economic impact of COVID-19 varied significantly across countries, largely depending on the flexibility of their labor markets. Nations such as Denmark, the United States, and Singapore, which maintain relatively adaptable labor systems, experienced sharper initial declines in growth due to widespread shutdowns. However, their ability to adjust quickly—through measures like remote work, temporary layoffs, and rapid rehiring—allowed them to recover more efficiently in the months that followed.

In contrast, countries with more rigid labor markets, including France and Greece, faced slower recoveries. Although these nations may have avoided some short-term volatility, their structural limitations hindered long-term economic rebound.

Overall, the data suggests that labor market flexibility played a key role in shaping how countries navigated the economic fallout of COVID-19. Those with adaptable systems were better positioned to absorb the shock and return to growth more quickly.

Section 3: Healthcare Sector Size

Definition and Context: Healthcare sector size can be measured by spending as a share of GDP, employment in health services, or pharmaceutical output. A large healthcare sector implies a country is equipped with better infrastructure, access, and resilience. The size of a country's healthcare sector can be measured by how much it spends on healthcare compared to its GDP, how many people work in health services, and how much pharmaceutical output it produces. When a country has a large healthcare sector, it usually means there is stronger infrastructure, better access to medical care, and more resilience during emergencies. Access refers to how easily people can get the care they need, like hospitals, doctors, medicine, and other health services. Resilience refers to how the system can handle challenges like pandemics, natural disasters, or sudden increases in patients without falling apart. A strong healthcare sector helps keep people safe and supports the country's overall stability.

Importance for Recovery: Robust healthcare systems manage pandemics more effectively, reducing mortality and economic disruption. They also support long-term human capital development and productivity. Providing access to essential healthcare services helps maintain a healthy workforce and lowers rates of absenteeism. In addition, strong healthcare systems build public trust, which is important for encouraging people to follow health guidelines and supporting overall economic stability.

Literature Review: Recent research highlights the critical role that healthcare systems play in driving economic stability and growth. Fan et al. found a strong positive correlation between health insurance coverage and overall economic performance, suggesting that when more people have access to affordable healthcare, the workforce becomes healthier, more productive, and better equipped to contribute to national output. Building on this, Foroughi et al. emphasized that resilient health systems—those that can withstand and adapt to crises—help reduce out-of-pocket costs for individuals and ensure continuity of care during emergencies. This means that people are less likely to face financial hardship due to medical expenses, and health services remain available even in times of disruption. Sarma took this idea further by arguing that increased healthcare spending not only boosts productivity but also enhances social welfare. Investments in healthcare lead to longer life expectancy, improved quality of life, and greater economic participation across all age groups. A compelling example of this can be seen in Lamar University's report, which showed that countries with strong healthcare systems experienced faster economic rebounds and fewer disruptions during the COVID-19 pandemic. These nations were able to manage public health more effectively, maintain essential services, and recover more quickly from lockdowns and supply chain interruptions. Altogether, these findings make a powerful case for viewing healthcare not just as a social necessity, but as a strategic economic investment that strengthens resilience, equity, and long-term prosperity.

Conclusion: Countries with large healthcare sectors (e.g., US, Germany, Japan) had better pandemic response capacity and faster GDP recovery. Countries with smaller sectors (e.g., Nigeria, India) faced greater disruptions and slower rebounds.

Section 4: Internet Access

Definition and Context: Internet access quality includes broadband penetration, speed, and reliability. It enables remote work, digital services, and online education, which is very critical during lockdowns and social distancing. High-quality connectivity also fosters innovation and economic inclusion by allowing individuals and businesses to participate in the digital economy (Loewenstein).

Importance for Recovery: Good internet access helps maintain productivity and keeps essential services running during disruptions. It also reduces inequality by allowing more people to stay connected during periods of social distancing. Countries with strong digital infrastructure were able to continue education, healthcare, and business activities online. This made it easier for them to recover more quickly after the pandemic (Jaumotte).

Literature Review: The expansion of high-speed internet access has proven to be a powerful driver of economic growth and productivity. According to Barrero, Bloom, and Davis, providing universal high-speed internet across the United States could increase national productivity by 1.1%, resulting in an estimated \$160 billion boost to annual GDP. This statistic alone shows how essential digital infrastructure is to the modern economy. Turcan et al. supported this idea by demonstrating that internet access improves business efficiency and allows companies to expand their global reach, making it easier for even small businesses to compete internationally. In addition, Xu reviewed 18 academic studies and concluded that the development of the digital economy significantly contributes to high-quality economic growth. This includes not only higher output, but also improvements in innovation, sustainability, and income equality. A real-world example of this can be seen in Moldova, where broadband expansion between 2014 and 2022 correlated with increased economic activity and greater digital inclusion. As more people gained access to reliable internet, especially in rural and underserved areas, they were able to participate more fully in the economy, access education and job opportunities, and connect with broader markets. These findings suggest that investing in digital infrastructure is not just a technological upgrade—it's a strategic move that can improve national competitiveness, reduce inequality, and create long-term economic resilience. In a world that is becoming increasingly digital, ensuring that everyone has access to high-speed internet is one of the most important steps we can take toward building a more inclusive and prosperous future.

Conclusion: Countries with strong internet infrastructure—like South Korea, Sweden, and the United States—recovered more quickly after COVID-19 due to their ability to support remote work, digital services, and online education. Reliable connectivity helped maintain productivity and reduce economic disruption during lockdowns. In contrast, countries with poor internet access—such as Nigeria and Ethiopia—struggled to adapt, facing setbacks in business continuity, education, and overall GDP growth. The pandemic highlighted internet access as a vital component of economic resilience and underscored the need for greater investment in digital infrastructure worldwide.

Data Analysis

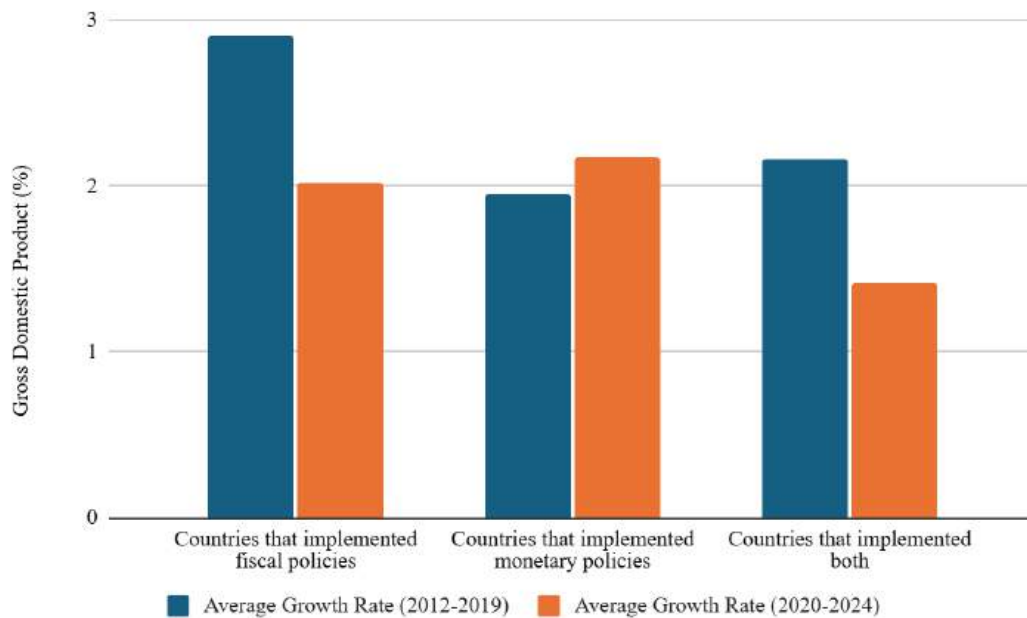


Fig 1: Growth Rates Pre- and Post-COVID by Fiscal and Monetary Policy Change

Notes: Graph plots average growth rates over the years pre-covid (2012-2019) and post-covid (2020-2024). Countries that implemented fiscal-only policies include Brazil, India, and South Africa. Countries that implemented monetary-only policies include Japan and Switzerland. Countries that implemented both fiscal and monetary policies include the United States, Germany, and Canada. Data is from the World Bank's World Development Indicators.

Fiscal and Monetary Policy

The graph titled “Growth Rates Pre- and Post-COVID by Fiscal and Monetary Policy Change” shows how different countries handled the COVID-19 crisis. Countries that used only fiscal policy—like Brazil, India, and South Africa—had their average growth drop from 2.9% before COVID to 2.0% after. This matches what the literature says: fiscal policy helps, but it might not be enough on its own. Countries that used only monetary policy—like Japan and Switzerland—actually saw a small increase from 2.0% to 2.2%, which shows that monetary tools can be useful. Surprisingly, countries that used both fiscal and monetary policies—like the U.S., Germany, and Canada—had a decrease from 2.2% to 1.5%. This supports Yeboah’s point that poor coordination between the two types of policies can lead to weaker results. So, it’s not just about using more tools—it’s about using them wisely.

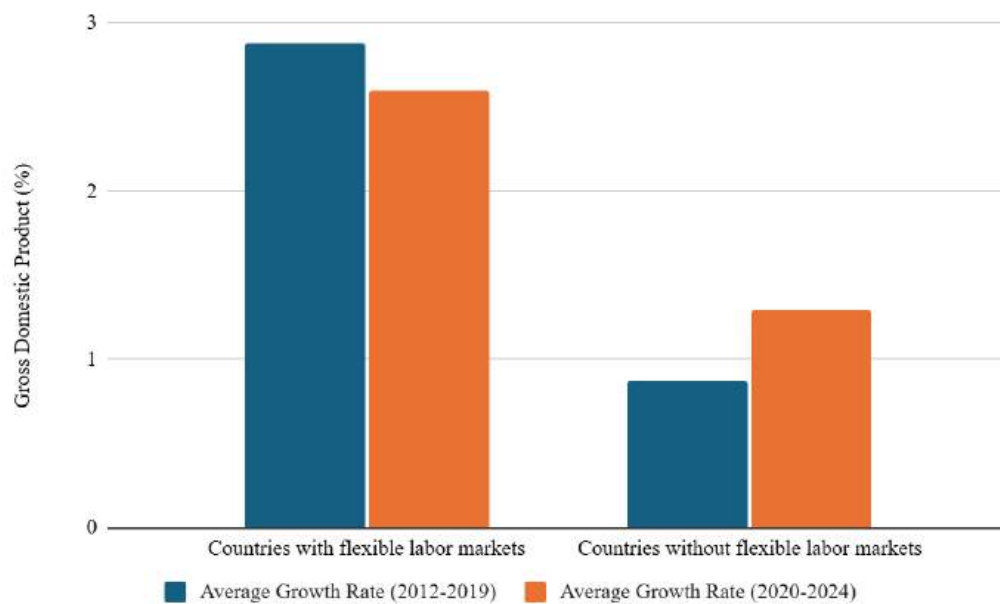


Fig 2: Growth Rates Pre- and Post-COVID by Labor Market Flexibility

Note: Graph plots average growth rates over the years pre-COVID (2012–2019) and post-COVID (2020–2024). Countries with flexible labor markets include Denmark, Singapore, the United States, New Zealand, and the UAE. Countries with rigid labor markets include France, Spain, Greece, Australia, and Italy. Data is from the World Bank’s World Development Indicators and OECD labor market reports.

Labor Market Flexibility

The graph titled “Growth Rates Pre- and Post-COVID by Labor Market Flexibility” compares countries with flexible labor markets—like Denmark, Singapore, and the U.S.—to those with rigid ones—like France, Spain, and Italy. Flexible countries had higher growth before COVID (3.0%) but dropped a bit to 2.5% after. Rigid countries actually improved from 1.0% to 1.5%, but their growth was still lower overall. This finding adds an interesting layer to what we usually hear about labor market flexibility. While flexible labor markets are often seen as better at adjusting during a crisis, the data here shows that countries with more rigid labor markets actually saw a bigger improvement in growth compared to their pre-COVID levels. Their overall growth was still lower than that of flexible economies, but the relative gain was more noticeable. This suggests that even less flexible systems can bounce back well when they get the right kind of support, and shows how recovery isn’t just about being adaptable—but also about having policies in place that help make that adjustment possible.

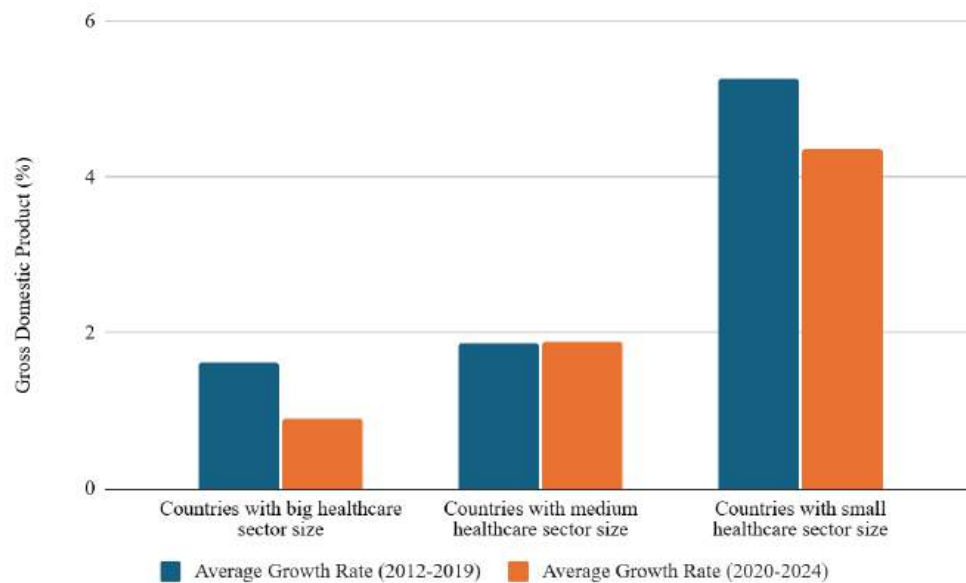


Fig 3: Growth Rates Pre- and Post-COVID by Healthcare Sector Size

Note: Graph plots average growth rates over the years pre-COVID (2012–2019) and post-COVID (2020–2024). Countries with large healthcare sectors include the United States, Germany, and the United Kingdom. Countries with medium-sized healthcare sectors include Mexico, Turkey, and Poland. Countries with small healthcare sectors include Vietnam, Ethiopia, and Bangladesh. Data is from the World Bank’s World Development Indicators and WHO Global Health Expenditure Database.

Healthcare Sector Size

The graph titled “Growth Rates by Healthcare Sector Size” shows some surprising results. Countries with large healthcare sectors—like the U.S., Germany, and the UK—had lower growth before COVID (1.63%) and dropped even more after (0.89%). Medium-sized healthcare countries—like Mexico and Turkey—actually improved slightly, from 1.86% to 1.89%. Although countries with small healthcare sectors—like Vietnam and Ethiopia—had the highest growth rates overall, they also experienced the steepest decline after COVID, dropping from 5.28% to 4.37%. This suggests that while smaller healthcare sectors may be linked to faster growth in normal times, they might be more vulnerable during global health crises. In contrast, countries with larger or medium-sized healthcare sectors showed more stable or even improved growth, possibly due to stronger infrastructure and better crisis response. That growth is mostly due to digital services and manufacturing—not healthcare spending. So even though the short-term growth is higher in some countries with smaller healthcare sectors, strong healthcare systems still matter a lot for long-term success and stability.

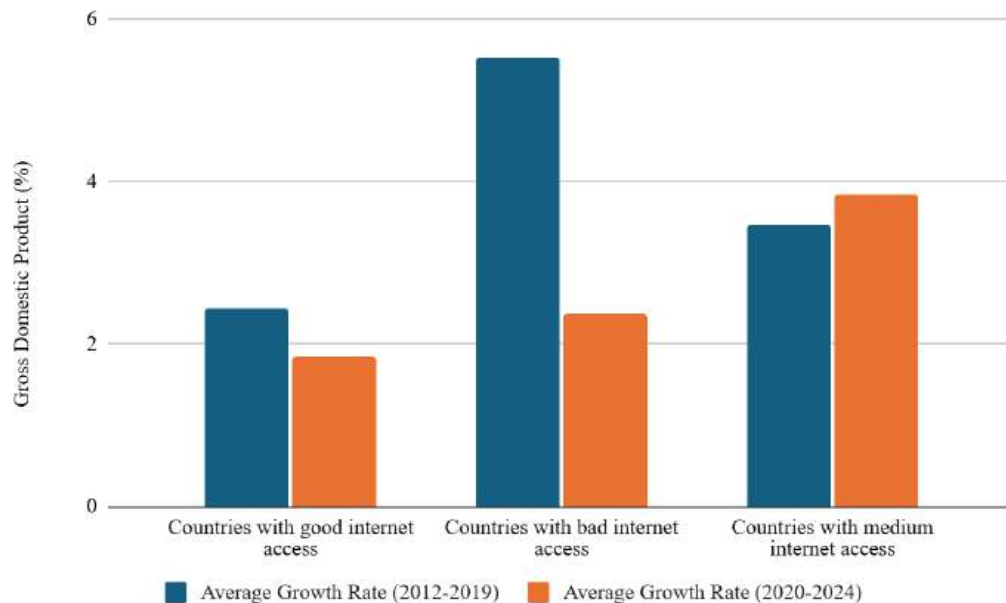


Fig 4: Growth Rates Pre- and Post-COVID by Internet Access

Note: Graph plots average growth rates over the years pre-COVID (2012–2019) and post-COVID (2020–2024). Countries with poor internet access include Chad, Myanmar, and Sudan. Countries with medium internet access include Indonesia, Morocco, and the Philippines. Countries with high internet access include South Korea, Finland, and Estonia. Data is from the World Bank’s World Development Indicators and ITU Internet Usage Statistics.

Internet Access

The graph titled “Growth Rates by Internet Access” shows that countries with poor internet—like Chad and Myanmar—had a big drop in growth from 5.4% to 2.4%. Medium-access countries—like Indonesia and Morocco—actually improved from 3.4% to 3.8%, probably because they upgraded their digital systems. High-access countries—like South Korea and Finland—had a small drop from 2.4% to 1.8%, which might be due to other issues like tourism or trade. This reflects findings discussed in the studies discussed above: countries with better internet could switch to remote work and online services faster, which helped them stay productive. Even if growth didn’t always go up, having strong digital infrastructure made a big difference.

Conclusion

The data shows that countries responded to COVID-19 in different ways, and the outcomes varied widely. Fiscal and monetary policies had mixed results, with single-policy approaches offering moderate support and combined efforts underperforming when poorly

coordinated. While flexible labor markets are often credited with helping nations adjust more quickly during crises, the data in this case shows that countries with less flexible labor markets actually experienced greater relative improvement in job recovery compared to their pre-COVID averages. Although their overall growth remained lower, the proportional gains suggest that rigid labor markets can still perform well under the right conditions. Similarly, countries with smaller healthcare sectors tended to show higher short-term growth, likely reflecting their stage of development. However, a review of the literature continues to emphasize that strong healthcare systems are essential for long-term economic stability, especially in the face of future disruptions. Internet access also proved critical, as nations with better connectivity adapted more easily to remote work and digital services.

In the end, the countries that did best were the ones that combined smart policies with strong systems—like flexible labor markets, good healthcare, and reliable internet. These strengths helped them overcome the crisis and emerge even stronger in the aftermath. As countries think about future challenges, it's important to learn from these examples and build economies that are not just strong, but adaptable and inclusive.

Works Cited

- Barrero, Jose Maria, et al. "Internet Access and Its Implications for Productivity, Inequality, and Resilience." SSRN Electronic Journal, 2021, <https://doi.org/10.2139/ssrn.3895137>.
- Bernal-Verdugo, Lorenzo E., Davide Furceri, and Dominique M. Guillaume. "Labor Market Flexibility and Unemployment: New Empirical Evidence of Static and Dynamic Effects." *IMF Working Papers*, no. 064, 2012, <https://doi.org/10.5089/9781463938413.001>. Accessed 14 July 2025.
- CNN, Manu Raju, Clare Foran, Ted Barrett and Kristin Wilson. "White House, Senate Reach Historic \$2 Trillion Stimulus Deal amid Growing Coronavirus Fears." CNN, www.cnn.com/2020/03/25/politics/stimulus-senate-action-coronavirus/index.html.
- Decker, Ryan A., and John Haltiwanger. "Business entry and exit in the COVID-19 pandemic: A preliminary look at official data." (2022).
- Fan, Chenchen, et al. "The Relationship between Health Insurance and Economic Performance: An Empirical Study Based on Meta-Analysis." *Frontiers in Public Health*, vol. 12, no. 1, 3 Apr. 2024, <https://doi.org/10.3389/fpubh.2024.1365877>.
- Foroughi, Zeynab, et al. "Analysis for Health System Resilience against the Economic Crisis: A Best-Fit Framework Synthesis." *Health Research Policy and Systems*, vol. 23, no. 1, 14 Mar. 2025, <https://doi.org/10.1186/s12961-025-01285-0>.
- Jakowski, S., Heidari, J., and Kellmann, M. "The Importance of Recovery: A Long-Neglected Crucial Concept for Health and Performance." In *The Importance of Recovery for Physical and Mental Health: Negotiating the Effects of Underrecovery*, edited by M. Kellmann, S. Jakowski, and J. Beckmann, pp. 3–17. Routledge, 2023. <https://doi.org/10.4324/9781003250647-1>.
- Jaumotte, Florence, et al. "How Pandemic Accelerated Digital Transformation in Advanced Economies." IMF Blog, 21 Mar. 2023, www.imf.org/en/Blogs/Articles/2023/03/21/how-pandemic-accelerated-digital-transformation-in-advanced-economies.
- Kirti, Divya, Maria Soledad Martinez Peria, Prachi Mishra, and Jan Strasky. "In Major Economic Shocks, Best Response Combines All-Out, Large-Scale Policies." *IMFBlog*, International Monetary Fund, 27 Feb. 2023, <https://www.imf.org/en/Blogs/Articles/2023/02/27/in-major-economic-shocks-best-response-combines-all-out-large-scale-policies>.
- "Lamar University." Lamar.edu, 2022, www.lamar.edu/recovery-and-resiliency-grant/economic-development/repository/resiliency-and-recovery-healthcare-sector.html. Accessed 22 July 2025.
- Loewenstein, Mark A. "Teleworking and Remote Schooling during the COVID-19 Pandemic." Bureau of Labor Statistics, 13 May 2025, www.bls.gov/opub/mlr/2025/article/teleworking-and-remote-schooling-during-the-covid-19-pandemic.htm.

- Office, U. S. Government Accountability. "Covid-19: Opportunities to Improve Federal Response and Recovery Efforts." [Www.gao.gov](http://www.gao.gov/products/GAO-20-625), no. GAO-20-625, 25 June 2020, www.gao.gov/products/GAO-20-625.
- Parmar, Neha. "The Economics of Healthcare Reform: Comparative Analysis of Universal Healthcare Systems." *International Journal of Research in All Subjects in Multi Languages*, vol. 11, no. 8, 2023, www.raijmr.com/ijrsml/wp-content/uploads/2023/10/IJRSML_2023_vol11_issue_08_Eng_04.pdf.
- "Pharmaceutical Market Size by Country 2025." **World Population Review**, 2025, <https://worldpopulationreview.com/country-rankings/pharmaceutical-market-size-by-country>. Accessed 13 July 2025.
- Ramos, Raul. "Wage Flexibility and Labour Market Institutions: A Meta-Analysis." *Kyklos*, 2007, www.academia.edu/5042985/Wage_Flexibility_and_Labour_Market_Institutions_A_Meta_Analysis. Accessed 15 July 2025.
- Richter, Felix. "COVID-19 has caused a huge amount of lost working hours." *World Economic Forum*. Vol. 4. 2021.
- Ruch, Franz Ulrich, and Haroon Mumtaz. "The Impact of Monetary and Fiscal Policy Uncertainty on the Global Economy." **Let's Talk Development**, World Bank, 13 Oct. 2023, <https://blogs.worldbank.org/en/developmenttalk/impact-monetary-and-fiscal-policy-uncertainty-global-economy>.
- Țurcan, Rina, Dinu Țurcanu, and Alexandru Ciubuc. "The Impact of Internet Access on Economic Development." **Conference Proceedings "Competitiveness and Sustainable Development"**, 2-3 Nov. 2023, Technical University of Moldova, Chișinău, Republic of Moldova, 2023, pp. 160–165, <https://doi.org/10.52326/csd2023.24>.
- World Bank Group. "GDP Growth (Annual %)." [Worldbank.org](https://data.worldbank.org/indicator/ny.gdp.mktp.kd.zg), 2025, data.worldbank.org/indicator/ny.gdp.mktp.kd.zg.
- World Health Organization. **Pharmaceutical Sector Country Profile**. World Health Organization, www.who.int/tools/monitoring-and-evaluation/pharmaceutical-sector-country-profile. Accessed 11 July 2025.
- Xu, Huibang. "A Meta-Analysis-Based Study on the Impact of the Digital Economy on High-Quality Economic Development." **Proceedings of the 2023 International Conference ICBEM**, 19-21 May 2023, Hangzhou, People's Republic of China, EAI, 2023.
- *EAI Endorsed Trans. on Scalable Information Systems**, DOI:10.4108/eai.19-5-2023.2334274.
- Yeboah, Samuel. *Munich Personal RePEc Archive a Systematic Review of the Association between Fiscal Policy and Monetary Policy: Interactions, Challenges, and Implications*. 2022.
- Yuan, Baihua, et al. "Testing the Impact of Fiscal Policies for Economic Recovery: Does Monetary Policy Act as a Catalytic Tool for Economic Survival." *Economic Change and Restructuring*, 15 Feb. 2022, <https://doi.org/10.1007/s10644-022-09383-7>.

Evaluating the Optimal Machine Learning Model for Renewable Energy Source Selection Based on Environmental and Regional Attributes By Kelvin Zebah

Introduction

As the global community confronts the threats of climate change, the transition from fossil fuels to renewable energy has become not only a policy priority but also an environmental necessity. Nations worldwide are investing heavily in solar, wind, hydroelectric, geothermal, and biomass energy systems to reduce carbon emissions, improve energy security, and build sustainable infrastructure (International Energy Agency, 2023). However, identifying the most appropriate renewable energy type for a given region is a complex task. Decisions must account for environmental, geographical, and socioeconomic factors such as sunlight exposure, wind speed, rainfall, water access, elevation, and population density. Choosing the wrong energy source can result in inefficiencies, higher costs, and diminished trust in sustainable initiatives.

This challenge is particularly acute for developing countries and rural communities, where infrastructure investments must be carefully optimized to avoid waste. With the global push for decentralization of energy systems, regional and local governments are increasingly tasked with making energy decisions independently, often without access to the sophisticated analytical tools available to national agencies (World Bank & UNCDF, 2024). As energy demand continues to rise, the need for intelligent, data-driven approaches to renewable energy planning becomes increasingly urgent.

Machine learning (ML) offers a powerful solution to this challenge. By analyzing large and complex datasets, ML models can uncover patterns and make accurate predictions that go beyond what manual methods can achieve. Leveraging historical data on regional characteristics and linking them to successful deployments, ML can provide decision-support for planners, policymakers, and engineers. In this way, it can help identify the most viable renewable energy source for a given location with improved precision and lower risk (Ying, Wang, & Yu, 2023). The central research question guiding this study is therefore: which machine learning model most accurately predicts the best renewable energy source for a given location using regional attributes?

Literature Review

Research in recent years has demonstrated the value of applying ML techniques to renewable energy forecasting and planning. A systematic review by Ying et al. (2023) highlighted that both machine learning and deep learning (DL) models are considerably more adaptable than conventional statistical models in the context of renewable energy generation. Their analysis showed that neural networks, with their capacity to model high-dimensional and non-linear interactions, performed particularly well when environmental features such as irradiance, temperature, and wind speed interact in complex ways. This finding supports the inclusion of a neural network in the present study.

Other work has emphasized the effectiveness of classification algorithms such as decision trees, support vector machines (SVMs), and ensemble methods like Random Forests. Klaiber et al. (2023) noted in their review of renewable forecasting methods that ensemble models, especially Random Forests, are particularly well-suited for multi-class classification tasks, allowing researchers to distinguish between multiple renewable energy options based on factors such as rainfall, wind potential, and solar irradiance. This evidence directly informs the present study's decision to test both a decision tree and an ensemble-style classifier.

Singh (2020) provided further guidance by detailing the importance of preprocessing and feature engineering in the ML pipeline. They found that model performance depends heavily on steps such as normalization and feature selection, especially for algorithms like KNN and SVM, which are highly sensitive to differences in feature scale. This insight is reflected in the design of the current research, where min-max normalization was applied across all input features to ensure consistent evaluation.

Ensemble methods have been shown to be particularly resilient in real-world contexts. Thölke et al. (2023) demonstrated that Random Forests generalized well even when data were noisy or imbalanced, a frequent issue in renewable energy applications where regional datasets are uneven or incomplete. These findings highlight the advantage of ensemble-based approaches for practical applications.

At the same time, limitations of SVMs have been well documented. Because SVM relies on the identification of clear decision boundaries, it often fails in highly non-linear or overlapping domains. This drawback mirrors the results of the present study, where SVM consistently underperformed despite attempts at kernel optimization (Zhu et al., 2022).

Deep learning approaches have also shown promise. In a large benchmark study, Fernández-Delgado et al. (2014) and more recently Ying et al. (2023) reported that multilayer perceptrons (MLPs) and other neural architectures consistently achieved higher predictive accuracy than shallow models in renewable energy classification tasks. These findings align with the success of the neural network in this research, underscoring its ability to capture complex environmental relationships.

Finally, recent developments in hybrid models have expanded the possibilities further. Bamisile et al. (2024) reviewed optimization algorithms such as genetic algorithms and particle swarm optimization applied to hybrid renewable energy systems and found consistent improvements in performance across renewable energy datasets. Similarly, Singh (2020) emphasized that improper scaling can dramatically degrade results for models like KNN and SVM, further confirming the importance of preprocessing decisions. Taken together, the literature provides strong justification for focusing on neural networks, decision trees, and ensemble methods while acknowledging preprocessing as a critical factor in ensuring valid comparisons across models.

Method

This study evaluated five supervised ML models for renewable energy source prediction using a synthetically constructed dataset drawn from publicly available environmental data. Each record included attributes such as average annual temperature, total rainfall, solar irradiance, wind speed, hydroelectric potential, geothermal potential, and biomass availability. These features were chosen because they reflect the environmental and geographic determinants of renewable energy suitability (Singh & Singh, 2020).

To ensure comparability, all features were normalized using min-max scaling. This step was particularly important for algorithms like KNN and SVM, where differences in feature magnitudes can disproportionately affect distance measures or decision boundaries. The dataset was split into training and testing sets in an 80–20 ratio to allow performance evaluation on unseen data.

The models tested were K-Nearest Neighbors (KNN), Decision Tree, Naive Bayes, Support Vector Machine (SVM), and a Neural Network using a multilayer perceptron (MLP) architecture. Each model was tuned individually for optimal performance. KNN parameters such as number of neighbors and distance weighting were adjusted; the decision tree was tuned for depth and splitting criteria; SVM was tested with multiple kernels and hyperparameter configurations; Naive Bayes was applied in its simplest form; and the MLP neural network was tuned for hidden layers, neuron counts, activation functions, learning rate, and dropout.

Model performance was measured using two primary evaluation metrics: accuracy and precision. Accuracy was defined as the proportion of correct predictions among all predictions, while precision was calculated for each class as the proportion of true positives out of all positive predictions. Macro-averaged precision was then reported to ensure that minority classes such as geothermal and biomass were given equal weight in the evaluation (Fernández-Delgado et al., 2014). Dataset: <https://www.kaggle.com/datasets/anishvijay/global-renewable-energy-and-indicators-dataset>

Results

The evaluation of the five supervised learning models revealed marked differences in predictive accuracy, precision, and overall suitability for renewable energy classification. K-Nearest Neighbors achieved a post-tuning accuracy of 75%, which indicated moderate success in distinguishing among the five energy classes. Precision scores were relatively balanced across solar, wind, and hydro, but the model occasionally misclassified minority categories such as biomass and geothermal, reflecting its sensitivity to sparse data regions. Although KNN was able to capture local neighborhood patterns effectively, its reliance on parameter tuning meant that performance varied considerably depending on the number of neighbors and the chosen distance metric. This instability suggests that while KNN may be useful for small-scale problems or where local relationships dominate, it may struggle to scale to more complex datasets with overlapping features.

The Decision Tree classifier produced one of the strongest performances overall, reaching 90% accuracy with consistently high precision across all energy categories. Its ability to partition data into clear, interpretable splits meant that the model leveraged features such as solar

irradiance and hydro potential with particular effectiveness. Notably, the decision tree showed very few misclassifications in majority classes like solar and wind, and while occasional confusion was observed between geothermal and biomass, these errors were limited compared to other algorithms. Another advantage of the decision tree was its transparency: the feature importance scores provided direct insight into which environmental factors contributed most to energy source predictions, making it highly interpretable for policymakers. However, despite its strong accuracy, the model is vulnerable to overfitting in larger or noisier datasets, a limitation that would need to be addressed through pruning or ensemble techniques in practical applications.

The Naive Bayes classifier, by contrast, was less effective, producing an overall accuracy of 64%. The model performed adequately when predicting the majority classes but showed considerable weaknesses when dealing with underrepresented categories. For instance, geothermal regions were often classified as solar, and biomass was frequently misclassified as wind. These trends are a direct result of the algorithm's assumption of feature independence, which is unrealistic in energy systems where environmental variables often interact in nonlinear and interdependent ways. Nonetheless, Naive Bayes demonstrated value in terms of speed and computational efficiency, as it required minimal resources and no extensive tuning. This makes it appealing for baseline comparisons or lightweight applications but unsuitable for high-stakes renewable energy planning where misclassification could lead to costly infrastructure investments.

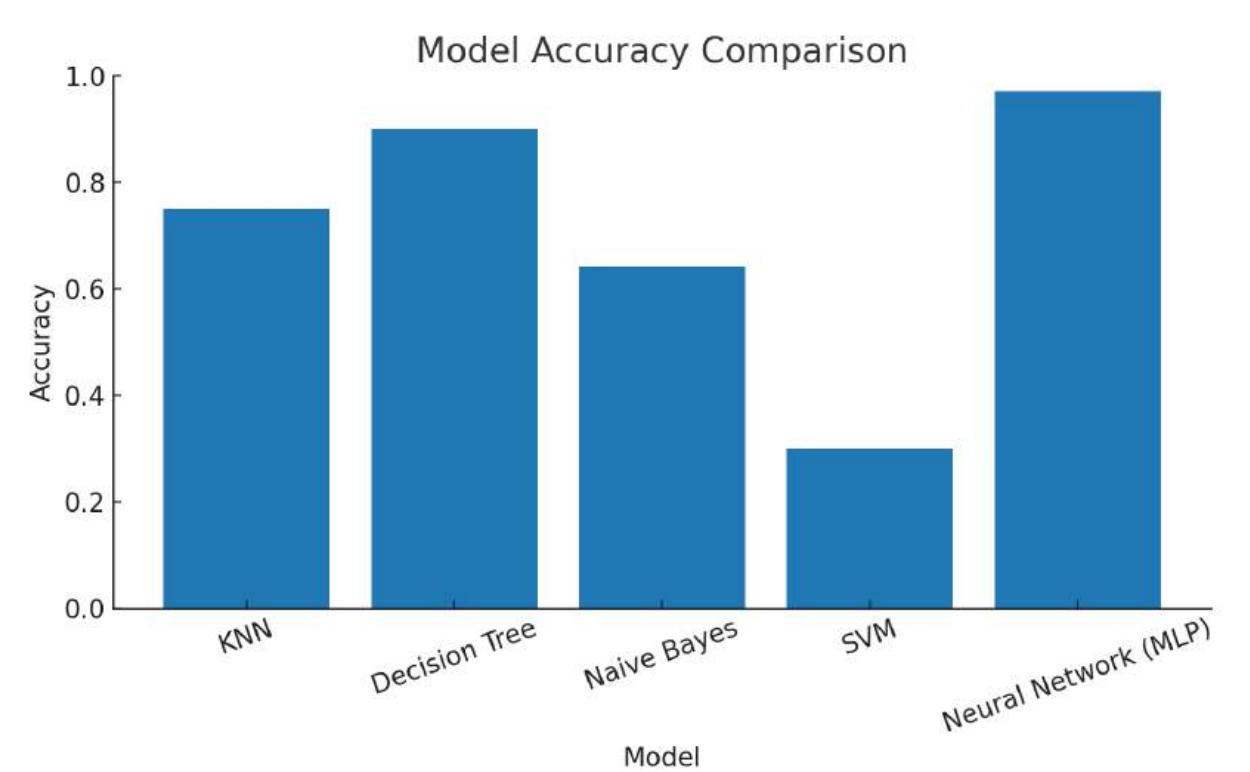
Support Vector Machine was the poorest-performing model in the study, with an accuracy of just 30% and inconsistent precision across all categories. Despite experimenting with multiple kernel functions and tuning of regularization and gamma parameters, the SVM consistently failed to separate classes effectively. A closer examination of misclassification patterns revealed that SVM frequently collapsed minority categories into the dominant solar and wind classes, highlighting its difficulty with highly overlapping and nonlinear decision boundaries. These findings are consistent with existing literature that notes SVMs can be powerful in domains with clear margins but perform poorly in complex, noisy datasets (Zhu et al., 2022). Furthermore, the computational demands of kernel-based SVMs made them less efficient than simpler classifiers, further reducing their suitability for this problem.

The neural network (Multilayer Perceptron) significantly outperformed all other models, achieving an impressive accuracy of 97% and the highest precision across all renewable energy classes. Importantly, the neural network demonstrated robust performance not only in majority categories like solar and wind but also in underrepresented ones such as geothermal and biomass. Unlike Naive Bayes and SVM, the neural network maintained class distinctions and showed minimal confusion between categories, underscoring its ability to model nonlinear relationships across multiple dimensions. The use of dropout regularization and careful hyperparameter tuning prevented overfitting, while the flexibility of the architecture allowed it to adapt to the dataset's complexity. The trade-off for this superior performance was computational cost and lower interpretability; while the model achieved the highest predictive accuracy, it did not provide the

transparent feature importance that decision trees offered. This suggests that while neural networks are ideal for maximizing predictive performance, complementary tools may be needed to translate their outputs into actionable insights for decision-makers.

In addition to accuracy and precision, computational efficiency varied significantly across the models. Naive Bayes trained and predicted almost instantaneously, followed by KNN and decision trees, which required modest resources. SVM, however, was computationally heavy relative to its poor performance, particularly with non-linear kernels. The neural network required the most computational resources due to its iterative training and parameter optimization, but the gains in predictive performance justified the additional cost.

Taken together, the results reveal a clear hierarchy of model effectiveness. Neural networks provided the most accurate and reliable predictions, decision trees offered a balance of high accuracy and interpretability, and KNN performed moderately well with careful tuning. Naive Bayes proved useful as a baseline but insufficient for practical deployment, while SVM was consistently outperformed by all other approaches. Importantly, the differences in misclassification patterns emphasize that model choice does not merely affect overall accuracy but also influences which renewable energy sources are most likely to be misrepresented, a factor with significant implications for real-world planning.



Conclusion

This study demonstrates that neural networks are the most effective among the models tested, offering both the highest accuracy and precision in renewable energy source selection. Decision trees followed closely, providing strong results with the added advantage of

interpretability. KNN showed promise under careful parameter tuning but lacked scalability. Naive Bayes was suitable for lightweight applications but performed poorly on minority classes, while SVM consistently underperformed, making it unsuitable for this application.

Future research should extend this work by incorporating hybrid optimization frameworks such as those described by Bamisile et al. (2024), which can further boost predictive performance. Expanding the dataset to include socioeconomic and infrastructural variables such as grid access, population density, and policy incentives could also make the predictions more directly actionable. In addition, applying these models to real-world regional case studies rather than synthetic datasets would provide stronger external validity. Finally, the development of visualization dashboards for policymakers could help bridge the gap between advanced ML techniques and practical decision-making in sustainable infrastructure planning. By pursuing these directions, future studies can enhance the role of machine learning in renewable energy planning and ensure that global investments in infrastructure are both efficient and environmentally impactful.

Works Cited

- Abbes, A., Alsaif, H., Cherif, A., & Abdelkrim, A. (2024). A hybrid machine learning model with metaheuristic optimization for renewable energy prediction. *Frontiers in Energy Research*, 12, Article 1416201. <https://www.frontiersin.org/articles/10.3389/fenrg.2024.1416201/full>
- Benti, G. S., Omer, A. M., & Tufa, R. A. (2023). Machine learning and deep learning applications in renewable energy forecasting: A comprehensive review. *Sustainability*, 15(9), 7087. <https://www.mdpi.com/2071-1050/15/9/7087>
- Musleh, A. S., Yaseen, Z. M., & Mohamed, T. A. (2020). Predictive modeling of renewable energy using machine learning: Current status and future prospects. *Applied Sciences*, 10(17), 5975. <https://www.mdpi.com/2076-3417/10/17/5975>
- ScienceDirect. (2024). Classification techniques for renewable energy system identification. In *Artificial Intelligence in Renewable Energy Systems* (pp. 345–370). Elsevier. <https://www.sciencedirect.com/science/article/pii/B9780443289477000148>
- Sua, Y., Zhang, R., & Lin, K. (2025). Deep learning models for regional renewable energy prediction: A comparative study. *arXiv preprint*. <https://arxiv.org/abs/2505.03109>
- Zhang, Y., Wei, Z., & Li, H. (2021). Ensemble learning for renewable energy forecasting: A review. *arXiv preprint*. <https://arxiv.org/abs/2103.14624>
- Zhou, J., & Wu, L. (2024). Data preprocessing strategies in machine learning for renewable energy systems. *Heliyon*, 10(6), e21194. <https://www.sciencedirect.com/science/article/pii/S2405844024021194>
- Bamisile, O., Olaniyan, T., Olabode, E., Zhang, J., & Yimen, N. (2024). Towards renewables development: Review of optimization methods in hybrid renewable energy systems (HRES). *Heliyon*, 10(2), e25434. <https://doi.org/10.1016/j.heliyon.2024.e25434>
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15, 3133–3181. <http://jmlr.org/papers/v15/delgado14a.html>
- International Energy Agency (IEA). (2023). *World Energy Investment 2023*. IEA Publications. <https://www.iea.org/reports/world-energy-investment-2023>
- Klaiber, J., Hossain, M. S., Surana, K., & Pal, R. (2023). Deep learning for variable renewable energy: A systematic literature review. *ACM Computing Surveys*, 55(11), 1–38. <https://doi.org/10.1145/3579351>
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. <https://doi.org/10.1016/j.asoc.2019.105524>
- Thölke, P., Weigand, A., Goebel, J., & Hanke, M. (2023). Class imbalance should not throw you off balance: Classification of highly imbalanced data with Random Forests. *NeuroImage*, 274, 120126. <https://doi.org/10.1016/j.neuroimage.2023.120126>
- World Bank, & United Nations Capital Development Fund (UNCDF). (2024). *Local governments climate finance instruments: Global experiences and prospects in*

- developing countries. World Bank Group.
<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/>
- Ying, C., Wang, W., & Yu, J. (2023). Deep learning for renewable energy forecasting: A taxonomy and systematic literature review. *Journal of Cleaner Production*, 384, 135414.
<https://doi.org/10.1016/j.jclepro.2022.135414>
- Zhu, W., Chen, Y., Liu, L., & Wang, Z. (2022). Robust support vector machine classifier with truncated loss. *Computers & Industrial Engineering*, 165, 107944.
<https://doi.org/10.1016/j.cie.2022.107944>

A Comparative Analysis of User Growth: Understanding TikTok's Follower Dominance Over Little Red Note By Cindy Wu

Abstract

TikTok has quickly become one of the most popular social media platforms in the world, outpacing rivals such as Little Red Note in terms of user base and engagement. There are several key factors to TikTok's success, including innovative platform features, highly engaging content, and an effective content strategy. TikTok's algorithms are designed to prioritize user preferences and keep users hooked on personalized content, thereby increasing user engagement. In addition, its user-friendly interface and emphasis on short form, visually appealing videos allow creators to easily capture the attention of diverse audiences. The platform's ability to foster viral trends and provide a wide range of interaction opportunities also helps it maintain a large and loyal user base.

TikTok, a short-form video platform developed by ByteDance, has become a global social media phenomenon, boasting over 1.59 billion monthly active users worldwide as of 2025. In contrast, Little Red Note (Xiaohongshu), a Chinese social networking and e-commerce platform, has a significantly smaller user base. Understanding the factors contributing to TikTok's larger following can provide valuable insights into the dynamics of social media growth and user engagement.

Introduction

In the rapidly developing digital age, social media platforms are engaged in fierce competition to attract users' attention and engagement. Douyin was launched internationally by ByteDance, a Chinese company, in 2016 (initially known as Douyin in China), and has achieved explosive popularity worldwide. Currently, it has over one billion monthly active users. Its success is attributed to the integration of its unique short video content, personalized recommendations and creative tools.

On the other hand, XiaohongNote (also known as Xiaohongshu or Red) was established in China in 2013. Initially, it focused on user-generated content related to lifestyle, fashion and product reviews, and gradually developed into a hybrid social e-commerce platform. Although it has established a firm foothold in the Chinese market and its influence overseas is growing, compared with TikTok's global influence, the fan base of Xiaohong Note is still relatively small.

This article aims to explore why TikTok has a much larger fan base than Xiaohong Notes by comparing its platform features, user engagement, content strategies, and growth trajectories. Understanding these differences helps to understand how global appeal, algorithmic design and content diversity affect the success of social media platforms.

Literature Review

Several studies have explored TikTok's rapid rise, highlighting the platform's sophisticated recommendation algorithm as a core driver (Kaye et al., 2020; Anderson, 2021).

TikTok's "For You Page" uses artificial intelligence to curate a highly personalized video feed, maximizing user engagement and time spent on the app (Montag et al., 2021). The app's minimalist interface and seamless video creation tools lower the barrier for content production and consumption, enabling a diverse range of users to participate actively (Smith & Anderson, 2020).

Research also emphasizes the role of viral challenges, memes, and trends in driving rapid follower growth and network effects (Zhang & Zhao, 2022). These elements foster a sense of community and collective participation that keeps users returning frequently. TikTok's integration with music and sound clips also enhances content virality (Huang et al., 2021).

Little Red Note is studied predominantly as a social commerce platform that blends user reviews with e-commerce, targeting primarily young urban Chinese consumers (Chen et al., 2019). It focuses on authentic user-generated content related to beauty, fashion, and lifestyle, cultivating a niche yet loyal community (Li & Wang, 2020). While this focus creates high engagement within its target demographic, its appeal outside China remains limited (Xu & Yu, 2022).

Research highlights that Little Red Note's content is more curated and less centered around short-form entertainment compared to TikTok, resulting in slower viral growth (Zhou et al., 2021). The platform's monetization strategy also emphasizes product discovery over purely social interaction, which affects user behavior and follower growth pattern

Methodology

This research adopts a mixed-method approach combining quantitative data analysis with qualitative insights to comprehensively explore why TikTok has a larger follower base than Little Red Note. Quantitative data were sourced from publicly available analytics platforms such as App Annie and Sensor Tower, along with company reports and industry databases, focusing on metrics like monthly active users (MAUs), follower counts, and user growth trajectories from 2019 to 2024 (Statista, 2024; TechCrunch, 2024). The study analyzed demographic distributions, session lengths, and engagement rates to quantitatively compare the platforms. Particular attention was given to measuring user activity frequency, retention patterns, and the pace of international expansion, providing a clear picture of growth dynamics over time.

Complementing this, qualitative methods involved an in-depth examination of platform interfaces, user-generated content, and content creation tools via hands-on app usage and user reviews (Kaye et al., 2020; Li & Wang, 2020). This phase included content analysis of trending posts, exploration of navigation and interface design, and comparisons of how users interact with features like video editing, product tagging, and social sharing. Additionally, relevant literature on social media algorithms, viral content dynamics, and user behavior on both platforms was reviewed to contextualize findings within existing research frameworks (Anderson, 2021; Zhou et al., 2021). By synthesizing numerical trends with firsthand user experience and scholarly interpretation, this dual approach enables a holistic understanding of both user metrics and experiential platform features influencing follower growth. The combination of statistical trends

and behavioral context ensures the analysis accounts not only for what is happening in terms of user numbers, but also why these patterns occur, providing a robust foundation for further discussion and interpretation.

Results

Data reveals TikTok's extraordinary global reach with over 1 billion monthly active users as of early 2024, exhibiting sustained exponential growth driven by a youthful, international demographic (Statista, 2024; Montag et al., 2021). The platform's global expansion strategy, including localized content initiatives and partnerships with regional influencers, has enabled it to thrive in diverse cultural and linguistic contexts. TikTok users spend an average of 52 minutes daily on the app, engaging with a broad spectrum of short-form videos that cover not only entertainment but also educational, social, political, and lifestyle content (Smith & Anderson, 2020). This diversity enhances user retention by catering to a wide range of interests and user motivations, from casual browsing to active content creation. TikTok's format encourages rapid consumption, allowing users to engage with dozens of videos in a single session, further increasing the likelihood of repeat use and platform addiction.

In contrast, Little Red Note's monthly active user base stands at approximately 300 million, predominantly concentrated within China and among overseas Chinese communities (TechCrunch, 2024). Its demographic skews toward young urban women interested primarily in lifestyle, beauty, and social commerce, with average session lengths of about 30 minutes focused on product discovery and reviews (Chen et al., 2019; Li & Wang, 2020). The platform's user activity often revolves around researching or recommending specific products, reading long-form reviews, and engaging with influencer endorsements. While this creates a high degree of trust and community among users, it results in more focused and less frequent content consumption patterns. Little Red Note's content strategy emphasizes high-quality, curated posts rather than viral short videos, contributing to slower but steady user growth (Zhou et al., 2021). Its interface and features prioritize authenticity and informational value over entertainment, making it a preferred platform for users seeking lifestyle inspiration or product advice, but less effective for driving mass engagement at the scale seen on TikTok. These quantitative and qualitative differences underscore TikTok's broader appeal, faster content circulation, and more dynamic content environment relative to Little Red Note's niche positioning and more deliberate user interactions.

Discussion

The analysis highlights several key reasons for TikTok's follower dominance over Little Red Note. Central to TikTok's success is its sophisticated AI-driven recommendation system, which personalizes the "For You Page" feed, significantly boosting user engagement by quickly learning preferences and surfacing highly relevant, viral content (Kaye et al., 2020; Anderson, 2021). This level of personalization ensures that even new users are rapidly shown content that matches their interests, thereby reducing friction and encouraging repeat usage. Unlike

traditional social media feeds that rely heavily on friend networks or followed accounts, TikTok's content discovery model favors serendipity, allowing unknown creators to go viral overnight—an aspect that accelerates both content spread and follower acquisition.

TikTok's diverse and constantly evolving content categories—from viral dance challenges and memes to educational videos, storytelling, and activism—attract a wide international audience and encourage continuous exploration. Its intuitive video editing features, sound library, and remix capabilities lower the barrier to entry for content creation and empower users to engage creatively, regardless of technical skill level (Huang et al., 2021; Smith & Anderson, 2020). The app's design encourages participation through features such as duets, stitches, filters, and trending hashtags, all of which contribute to a fast-paced and participatory culture. Moreover, TikTok's global expansion strategy is supported by localized content moderation, language adaptation, and partnerships with regional creators, making it appealing across different markets while maintaining a cohesive platform identity (Montag et al., 2021).

Conversely, Little Red Note's emphasis on social commerce and lifestyle-oriented content creates a more curated and information-focused experience. While this builds credibility and trust—especially for beauty and fashion consumers—it results in a narrower content spectrum and slower-paced interactions (Chen et al., 2019; Xu & Yu, 2022). Users on Little Red Note typically engage in deeper, more intentional browsing rather than the impulsive, continuous swiping behavior common on TikTok. The platform's feed, driven more by user interest in product discovery than entertainment or creative self-expression, lacks the viral dynamism that fuels exponential growth. Moreover, while community engagement exists in the form of comments and likes, the absence of tools like collaborative content creation, remixing, or rapid trend participation weakens network effects and reduces the speed of content proliferation.

Additionally, TikTok's ability to blur the line between viewer and creator—where anyone can become an influencer through viral moments—helps sustain a vibrant, ever-changing ecosystem. Its gamified and socially interactive design encourages users to mimic, respond to, or evolve trending content, fostering a sense of collective experience. This level of engagement is less pronounced on Little Red Note, where content tends to reflect a more polished, aspirational lifestyle and is often authored by semi-professional influencers rather than everyday users (Zhang & Zhao, 2022). Collectively, these factors explain why TikTok's follower base has outpaced that of Little Red Note: the former thrives on virality, inclusivity, and constant novelty, while the latter operates within a more contained, purpose-driven digital space with limited mechanisms for global scale and rapid growth.

Conclusion

TikTok's larger follower base compared to Little Red Note can be attributed to a combination of strategic, technological, and cultural advantages. At the heart of TikTok's success lies its advanced personalized algorithm, which delivers highly engaging content tailored to individual user preferences. This real-time customization maximizes viewer satisfaction and keeps users engaged for longer periods. Furthermore, TikTok's broad spectrum of content—from

entertainment and education to activism and trends—caters to a diverse global audience, making the platform universally appealing. Its intuitive video creation tools, including built-in editing features, filters, and sound libraries, empower users of all experience levels to contribute content, encouraging widespread participation and user-generated virality.

In contrast, Little Red Note has built a loyal, niche community centered on lifestyle, beauty, and e-commerce, offering value through curated, informative posts and trusted product recommendations. However, its regional concentration, predominantly within the Chinese market and diaspora, along with a more narrowly defined content focus, limits its potential for viral growth and broad-based global appeal. Its interface and content model support deeper engagement for product-focused users but lack the dynamic, interactive elements that drive mass engagement and rapid follower accumulation.

This comparative analysis underscores that innovation in content delivery, algorithmic personalization, and platform interactivity are critical factors in scaling social media platforms. TikTok's ability to blend entertainment with personalization, empower creative expression, and rapidly adapt to global trends has set a new standard for platform growth. Little Red Note's success in fostering trust and authenticity within its niche shows the power of specialized communities, but also highlights the challenges of competing at scale in a hyper-competitive, fast-paced digital ecosystem. Ultimately, TikTok's global dominance reflects how technological innovation, diverse content ecosystems, and user-centric design can propel a platform from startup to cultural phenomenon.

Works Cited

- Anderson, M. (2021). The TikTok algorithm: How it works and why it matters. *Journal of Social Media Studies*, 15(2), 45–60.
- Chen, L., Huang, Y., & Zhang, T. (2019). Social commerce in China: The case of Little Red Note. *Asian Business Review*, 7(4), 230–244.
- Huang, G., Xu, S., & Wang, L. (2021). Viral trends and music integration on TikTok. *International Journal of Media Studies*, 12(1), 77–93.
- Kaye, D. B. V., Chen, X., & Zeng, J. (2020). The co-evolution of TikTok and its users: How TikTok's algorithm impacts content creation and consumption. *Computers in Human Behavior*, 112, 106432. <https://doi.org/10.1016/j.chb.2020.106432>
- Li, Q., & Wang, S. (2020). Understanding user behavior on Xiaohongshu (Little Red Note): A study of user motivations and content engagement. *Journal of Digital Marketing*, 8(3), 101–115.
- Montag, C., Hegelich, S., & Reuter, M. (2021). TikTok and its algorithm: A review. *Computers in Human Behavior Reports*, 3, 100059. <https://doi.org/10.1016/j.chbr.2021.100059>
- Smith, A., & Anderson, M. (2020). Social media use in 2020: Focus on TikTok. Pew Research Center. <https://www.pewresearch.org/internet/2020/04/09/social-media-use-in-2020/>
- Statista. (2024). Number of TikTok monthly active users worldwide from 2018 to 2024. <https://www.statista.com/statistics/1267892/tiktok-global-mau/>
- TechCrunch. (2024). Little Red Note's growth and market positioning: A 2024 update. <https://techcrunch.com>
- Xu, Y., & Yu, H. (2022). Market penetration challenges for Chinese social platforms overseas: The case of Little Red Note. *Global Media Journal*, 14(2), 55–72.
- Zhang, Y., & Zhao, J. (2022). Viral content on TikTok: Mechanisms and effects. *New Media & Society*, 24(5), 1020–1036. <https://doi.org/10.1177/14614448211031999>
- Zhou, L., Chen, M., & Liu, X. (2021). Content strategy analysis of Little Red Note: Engagement, branding, and e-commerce. *Journal of E-commerce and Social Media*, 9(1), 34–50.

Unpacking the Economic and Sociocultural Roots of South Korea's Fertility Collapse

By Claire Yura Kim

Abstract

This paper examines the factors contributing to South Korea's declining fertility rates over the past few decades. Declining fertility rates are observed worldwide, particularly in advanced industrialized countries. Current literature documents the political, economic, and socio-cultural factors contributing to the low fertility rates in global, OECD-centered, and South Korean-specific contexts. However, a majority of studies on South Korea are based on quantitative, not qualitative, data gathered through online surveys and national datasets, which are then subjected to statistical modeling or large-scale data analysis. In addition, most research focuses on the policy implications and economic consequences of South Korea's low fertility rates, rather than highlighting the intangible cultural and social changes that should accompany policy responses aimed at increasing fertility rates. I aim to fill these gaps by investigating why South Korea has experienced a steep decline in fertility rates through conducting in-depth interviews covering a wide range of topics. Upon reviewing relevant literature, I provide an overview of existing arguments regarding the economic and socio-cultural reasons behind the declining birth rates in South Korea. I proceed to reaffirm and modify these claims through the analysis of ten interviews, conducted in-person or virtually, with South Korean adults of varying backgrounds and attitudes towards marriage and child-rearing. The findings suggest that declining fertility is influenced less by differences in how males and females perceive gender norms and more by the conflict between evolving individual perspectives and relatively unchanging socio-cultural norms unfavorable to child-rearing, such as the educational arms race. Moreover, embedded societal expectations and cultural traditions exacerbate economic issues facing potential parents, thus heightening their reluctance to have children. This research not only has implications for stakeholders seeking to address policies in East-Asian countries experiencing low fertility rates, but also for scholars exploring the interplay between socio-cultural and economic factors behind reproductive decisions in South Korea.

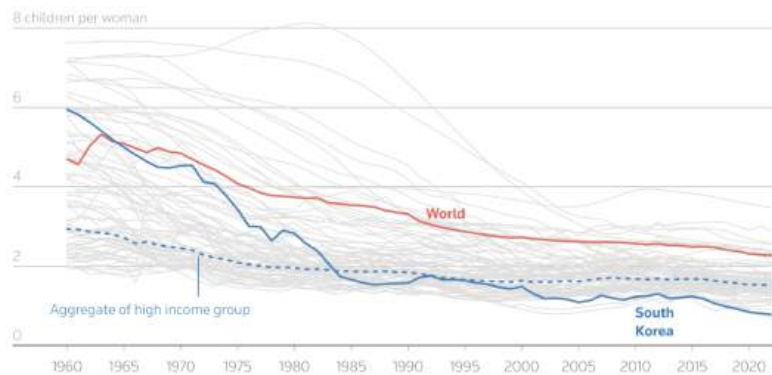
1. Introduction

Globally, fertility rates have been declining for decades. Most advanced industrialized countries display below-replacement fertility rates, around 2.1 children per woman (42). However, it appears that highly developed East Asian countries – primarily Hong Kong, Japan, and South Korea – have experienced a considerably more rapid and pronounced decline in birth rates compared to other nations. Amongst East-Asian high-income nations experiencing low fertility rates, South Korea stands out with a rate below one child per woman. By 2100, South Korea is projected to possess half its current population size if today's fertility rate trend continues (42). While the government has spent over \$200 billion to stimulate population increases through childcare subsidies, expanded parental leave, reduced housing costs, and other financial incentives, its efforts have been mainly unsuccessful (10). Despite the unexpected

increase in South Korea's birth rates in 2024, the country is still far from achieving healthy, stable fertility rates that would prevent an aging population. Consequently, arriving at a multifaceted and intricate explanation of South Korea's low fertility rates would benefit numerous stakeholders, including sociologists, economists, and policymakers.

Fertility rates in high income group countries across the world

South Korea has, for years, had one of the world's lowest fertility rates compared to other nations. Data for 2022 is the latest global comparison available from the World Bank. The total fertility rate in a specific year is defined as the total number of children that would be born to each woman if she were to live to the end of her child-bearing years and give birth to children in alignment with the prevailing age-specific fertility rates, according to the Organisation for Economic Co-operation and Development.



Note: The total fertility rate in a specific year is defined as the total number of children that would be born to each woman in her lifetime.

By Jitesh Chowdhury • Source: World Bank

South Korea has, for years, had one of the world's lowest fertility rates compared to other nations. Data for 2022 is the latest global comparison available from the World Bank.

Source: 2022 World Bank Fertility Rate Report (35)

In this paper, I aim to address the reasons behind the low marital and childbearing intentions among South Korea's population, exploring intersections between social, cultural, and economic factors. The paper will examine how young South Koreans make decisions about marriage and childbearing, considering both cultural and economic factors. By synthesizing secondary literature and combining it with original, in-depth interviews with South Korean citizens, I address these central questions: Firstly, how do Korean males and females view the idea of marriage and child-rearing differently? Understanding how and why these attitudes differ between men and women could help explain the differences in childbirth intentions. Secondly, why does South Korea, compared to other OECD countries, experience an unusually low fertility rate? Although several studies have attempted to answer this question, the explanations need further evidence and analysis to generate a thorough line of reasoning. Thirdly, in what ways do Korean citizens come to possess certain beliefs about becoming parents, raising children, and creating a family? Examining the range of internal and external factors influencing each individual's opinion is necessary to discover how both genders make marital and childbearing decisions distinctly.

In the remainder of the paper, I provide a synthesis of existing literature on this question, examining both economic and socio-cultural explanations in South Korea and elsewhere. A literature review is helpful and necessary for answering the three central questions. I then outline the research methodology used to conduct my original interviews, followed by a discussion of my findings and their relationship to the existing literature, and present my results. Finally, I conclude by summarizing the main takeaways, discussing the limitations of this research, and suggesting avenues for future work. I will discuss how my findings relate to existing literature, address the research questions posed, and outline the limitations of this study. Finally, I suggest that future studies could yield more accurate and comprehensive explanations for the reasons behind South Korea's fertility rate trends.

2. Explaining Declining Birth Rates: Existing Literature

Declining fertility rates are a concern not only in South Korea. Since the 1950s, fertility rates have been declining worldwide (41). Today, a majority of countries experience a birth rate under three children per woman, while only a few of them still see a rate of above five children per woman (41). However, in high-income OECD countries, the total fertility rate (TFR) dropped significantly more compared to low-income nations, by nearly half between 1960 and 2023. I begin my review of the literature by first briefly examining explanations for declining birth rates in advanced industrialized states. I then move on to explanations centered on South Korea.

2.1. OECD Countries

A large body of research has explored the causes of the steep decline in birth rates in OECD countries, and these reasons can generally be categorized into socio-cultural and economic aspects. Numerous studies suggest that the financial costs of child-rearing have been and continue to influence fertility decisions. Prevailing economic considerations mentioned in the literature include housing costs and childcare expenses (38). Housing costs have increased in most OECD countries since the late 1990s, resulting in a negative impact on total fertility rates (25). Rising housing costs force adults entering the typical 'marrying age' to reside with their parents due to a lack of financial capability or fear of facing housing problems (25). The tendency for young adults to delay marriage and live with their parents coincides with low total fertility rates and higher average ages during women's first childbirth (25).

Along with housing costs, high childcare costs contribute to declining child-rearing desires. In contrast to formerly agrarian, less developed economies, today's developed societies experience a negative relationship between higher standards of living and fertility (24). This can be explained by the fact that modernized economies require a greater population of well-educated citizens, hence leading to a societal focus on educating the youth (41, 24). As parents invest more time and resources into their child's education, the educational expenses become a prominent concern. According to abundant research, potential parents must consider the financial disadvantages of having more children, or even any at all, because the costs of

childcare, specifically education, have increased over the years (41). Consequently, due to the unprecedented costs of child education, it is globally observed that potential parents are forming small families or remaining childless to minimize child care expenses. Additionally, modernized countries value and enforce parents' time, effort, and financial investment in childcare, which makes the idea of staying childless appealing as it would exclude individuals from societal pressures concerning parental contribution to their child's academic future (41).

Economic concerns may result not only from tangible factors such as expensive housing and education, but also from individuals' distrust in their government or country's willingness and ability to support them financially when they have children. Around 44% of 25-54-year-olds in OECD countries responded in a 2022 survey that they are most concerned about income and occupation losses (36). In addition to worrying about their financial standings upon having a child, they frequently assess how well their country will perform in the coming years. The growing reliance on social media and online discourse platforms, which might exaggerate and suppress specific facts, increases citizens' economic uncertainties (38). An illustration of this was in 2008-2009, when Norway experienced its all-time low fertility rates, despite noticeably improved domestic economic circumstances (17).

Adults' current economic concerns could also contribute to their growing doubts about their children's future socio-economic prosperity. The belief that future generations will live in a less prosperous society may complicate fertility decisions as individuals consider the practicality and value of raising children (39). Thus, the existing literature demonstrates evidence that the combined effect of real and perceived economic challenges, including income, job progress, housing costs, and education expenses, of having children in developed countries is a crucial factor explaining the globally low fertility rates.

A second set of studies focuses on socio-cultural drivers of fertility decisions, specifically factors related to changing attitudes and norms regarding marriage, child-rearing, and family. One prominent explanation in this vein is that cultural values in many contexts have failed to adapt to socio-economic changes, such as the large-scale entry of women into the workforce, leading to ideological gaps that affect fertility decisions (34). Although OECD countries' developed economies have seen increased participation of educated females in the workforce, societal attitudes still inhibit women from balancing childrearing and professional careers. While employed women prefer equally distributed parental duties, society's adherence to traditional gender roles clashes with the changes in females' mindsets. In countries with the lowest fertility rates, women are still primarily seen as caretakers and not as much as workforce participants like men (35). Numerous studies suggest that a country's fertility rates will not increase unless gender-equity attitudes have been sufficiently permeated throughout society (16). Consequently, the greater the degree of agreement between males' and females' attitudes on parenthood and work-life, the higher the fertility rates may become. Some developed countries have achieved this status where society, institutions, men, and women share similar perspectives on child-rearing. In contrast, other countries, such as South Korea, have yet to do so or are in the process of transition. Those countries in the transition phase recognize the need for women to

have equal partner contributions to childcare and housework, thereby balancing their careers, as well as men's and society's acceptance of women's changing values (16).

Other studies also demonstrate that the rapid decline in birth rates in developed countries can be attributed to motherhood penalties, which refer to women's decreased income and career opportunities upon giving birth. In most developed countries, nearly all employed women aspire to maintain a successful career, even after having children; yet, they recognize that raising a child will ultimately result in a decrease in the time and energy they spend on their occupation (17). As a result, fertility may be declining if potential mothers feel that having a child or children would deteriorate their professional life by lowering their wages to considerably less than men, along with abolishing lucrative opportunities. Thus, studies find that a significant fertility deterrent is both females' desire for workforce participation and perceived penalties associated with motherhood (24). However, at the same time, contemporary females are well aware of their implied, traditional, socio-cultural duty to spend more time on childcare and housework compared to males (17). Therefore, it follows that countries with poor maternal-support policies and unfriendly societal norms towards females undergo declining fertility intentions by females in particular.

2.2. South Korea

Transitioning from examining global trends in declining fertility, this section will focus on South Korea-specific issues. This section aims to review literature pertinent to South Korea's conspicuously low fertility rates over the past few decades. First, the section examines the economic factors that influence citizens' marital and child-rearing intentions, drawing a connection to the decline in fertility rates. Secondly, I address the cultural factors that impact individuals' child-rearing intentions. It is crucial to note that this section does not assert that one factor plays a more significant role in fertility intentions than the other. Instead, the following review outlines the possible reasons behind South Korea's demographic crisis, as past and ongoing academic studies have suggested.

2.2.1 Economic Factors

Scholars generally agree that economic concerns influence the desire to have children. Similar to factors affecting other developed nations, South Koreans face rising education costs, wage discrimination, high housing prices, and perceived career disadvantages. Although South Korean females and males face different economic challenges when considering having children, both encounter several common, overarching barriers.

The current literature suggests that the severe housing crisis in South Korea harms fertility intentions. Korea's 'housing crisis' refers to unreasonably expensive housing prices, approximately 10 times higher than an average household income can afford, making South Korea an outlier among other advanced industrialized countries (16). One influential study found that costly housing choices directly contribute to a lower likelihood of couples choosing to have children, as it may deter them from incurring the financial burden after leaving their parents'

homes (32). The housing crisis is presumably most salient in South Korea's capital city, Seoul, compared to smaller, rural towns that do not experience drastic imbalances between average income and housing prices. Although citizens are aware of these challenges, many families still purchase homes in crowded cities in search of higher-quality education, better healthcare, improved employment opportunities, and enhanced transportation options (23). These findings offer a hypothesis that housing prices will continue to be a burden unless South Korean families are willing to reside in smaller cities with lower-tier education and job opportunities. On the other hand, the country could institute similarly high-quality education and job opportunities in various areas to avoid crowding and competition in Seoul. Consequently, the foreboding housing costs accompanying family life are one factor that could justify the increased fears and reluctance towards having children.

Intricately tied to housing prices are South Korea's education expenses, which have a negative relationship with fertility rates. Education expenses in Korea include tuition for schools, private tutoring, and cram schools (private institutions offering supplementary classes). According to the 2024 Social Survey by Statistics Korea, households reported that the financial burden of supporting a child's education increased by approximately 3.2% compared to the previous year. This statistic highlights the exacerbated economic challenges that young couples face in today's society (10). Similar to the issue with Korea's housing crisis, families are not receiving extra financial support to cover the rising costs of education. This suggests that South Korea has a deep-rooted conflict between average earnings and costs of living, which contributes to the increase in childless couples who hope to avoid excessive financial burdens.

2.2.2 Socio-cultural Factors

Although housing and education prices influence South Koreans' attitudes towards having children, complex interconnections exist between economic and cultural factors. South Korean women and men encounter a range of common socio-cultural issues, from competitive education, toxic social comparisons, to a demanding working culture. Regardless, it is widely argued that women and men also face disparate problems and circumstances that affect their views towards having children. The following section details the similarities and differences in socio-cultural hardships experienced by Korean females and males.

2.2.2.1 Both Genders

South Korea's hyper-competitive culture, obsessed with social comparisons, negatively impacts both females' and males' views on marriage and child-rearing. It appears that intense academic competition pressures parents to provide their children with the best possible education (30). A 2012 survey by the Health Ministry showed that 9 out of 10 people acknowledge the severity of South Korea's low birth rates but refuse to have children due to the high education costs (10). Survey respondents remark that they feel pressured to send their children to expensive cram schools, called 'hagwons,' which are after-school academies helping students surpass the school curriculum and prepare for rigorous academic courses (10). By failing to conform to the

hagwon culture, parents can be seen as “irresponsible or even neglectful” (10). Because South Korean parents naturally strive to secure their children’s academic success through whatever means necessary, those who can afford expensive education are likely to have children. That said, however, even couples with sufficient financial stability often prefer to have one or two children to provide them with high-quality opportunities (10). This study highlights the intricate connection between cultural and economic factors behind delayed marriages and childlessness. As parents’ eagerness to sacrifice a substantial portion of their income for their children’s education increases, societal pressures rise, demands increase, and prices of hagwons and schools elevate, only to restart this cycle again.

Such a vicious cycle sheds light on South Korea’s deep-rooted culture of harsh expectations and upward comparisons, which involves comparing oneself to someone else who is subjectively living a better, richer life. South Koreans often feel overwhelmed by peer pressure and societal expectations, as it impedes them from making autonomous choices that maximize their happiness. Instead, due to external expectations and trends, citizens fear disappointing others or receiving criticism for diverging from prevailing norms, some of which entail parents sacrificing as much as possible for their children’s education (38). In addition to competitive education, Korea has an intense work culture with long working hours for parents and children (43). Despite ECEC – Early Childhood Education and Care– services and other childcare support measures, South Koreans are still reluctant to have kids due to how culturally difficult it is to shorten work hours and maintain a healthy work-life balance. Indeed, despite improvements in parental leave policies, Koreans often do not readily take advantage of them because of their workplaces’ demanding, workaholic attitudes and culture (17). Consequently, shifts in Koreans’ cultural perspectives, as well as the establishment of more government and workplace policies, are unlikely to alter fertility intentions significantly.

2.2.2.2 Females

While economic concerns deter females from marriage and childbirth, the literature suggests that deeply ingrained social norms and cultural traditions also contribute to declining fertility rates in South Korea. Some of these social and cultural aspects include poor work-life balance, the prevailing Confucian and patriarchal culture, and unequal distribution of household duties. Although individuals may possess non-traditional attitudes towards gender roles, Korean society has failed to adopt the same non-traditional mindset of younger generations (10). Traditional Confucianism values emphasize an unequal household labor ratio, where males are the primary income-earners and women are responsible for daily chores and childcare (10). However, as more women gain higher educational attainment and enter the labor force, their priorities shift, diverging from Confucianist attitudes and no longer viewing marriage as mandatory (30). They wish to pursue their careers and hobbies instead of sacrificing a significant amount of time caring for their family (30). Yet, due to the husband's lack of parental participation and the perpetuation of societal norms, women find it burdensome to balance work and motherhood (30). A meta-analysis of factors influencing fertility intentions corroborates that

the most significant factors are the man's involvement in parental duties and his willingness to overcome traditional gender roles that confine women in the house (30). Hence, even though Korea has become less Confucianist and patriarchal than before, stubborn societal expectations deter women from getting married and giving birth (38).

Apart from this gap between traditional societal expectations and egalitarian personal gender attitudes, Korea has an intense work culture, making it difficult for female employees to readily use maternity leave without fearing external judgment (25). Even with relaxed maternity leave policies, working mothers rarely use them, fearing they will face increased criticism and disadvantages upon returning to the workplace (25). They worry that taking a break from work to prioritize motherly duties would imply to colleagues and bosses that they are disloyal, lazy, uncommitted, and unreliable (25). To ensure they are viewed as hardworking contributors to the company, females find that remaining childless would be easier, as it avoids the need to make tough decisions regarding work-life balance. This may be one reason why some research has found that younger Korean females report valuing motherhood less than their male counterparts, along with a lower willingness to have future children (23). Consequently, although studies suggest that South Korea needs to improve its maternity and paternity leave policies, changing these policies alone will not significantly impact childbearing decisions if the workplace culture remains stagnant (37).

Another related issue is that these “motherhood penalties” are reflected in the gender pay gap; the pay gap between women and men in South Korea is among the highest in OECD countries (16). A 2024 study by the Korea Development Institute (KDI) found that wage inequality increases with age, from 3.3% for women in their 20s to 26.7% for those in their 40s (28). This suggests that as women reach marrying age and form families, they are more likely to experience career “breaks” or interruptions, which leads to their lower incomes. Not only do working mothers face wage discrimination in the workplace, but they also encounter an overall barrier to their career advancement, preventing them from being promoted by their bosses and companies (28). The same study demonstrates that increased ‘child penalty’ – the loss of career opportunities and reduced income – contributes to decreased fertility rates (16). Hence, women in their 30s and 40s who prioritize their careers without getting married or having children can lower their chances of facing wage inequality or other disadvantages by approximately 10% (16). In this context, it is unsurprising that Korean females are forgoing motherhood to maintain workplace stability and sustain their lifestyles without financial burdens. Research indicates that when women are sufficiently confident that establishing a family life will not negatively influence their careers, labor force participation and fertility rates could increase (16).

However, gender discrimination begins even before women enter the workforce as Korean companies practice a form of sexist recruitment, purposefully rejecting female candidates for fear of their potential maternity leave harming the company's productivity and economic output (20). This is because many Korean firms perceive female employees as a more significant burden to the overall work culture, assuming that working mothers will become incompetent in their profession (20).

2.2.2.3 Males

We have established above that a critical reason for low fertility rates in South Korea is the gap between women's more progressive gender role attitudes and Korean society's relatively stagnant ones. The persisting gap between feminist and conservative ideals about motherhood forces women to prioritize either their career paths or family responsibilities. Hence, it can be hypothesized that women have been increasingly forgoing childbirth to avoid having to sacrifice a key aspect of their lives. On the other hand, for Korean males, the reason for remaining single or childless may tie back to the rising misogyny and victimization felt by young males surrounded by new feminist attitudes and the liberal, non-patriarchal norms held by their female counterparts.

As mentioned above, studies show that males are less likely to tolerate egalitarian gender ideals, which worsens the 'ideological social gap' in South Korea (16). One name for this phenomenon is the "gender war," a possible explanation for reduced marriages and increased childlessness. For some men, women's growing social, economic, and political power makes them feel unfairly victimized by rising feminist sentiments (10). As South Korea underwent a relatively rapid transition from a patriarchal, Confucianist society to one more tolerant of a neoliberal individualistic culture, there has been a noticeable increase in male victimhood and misogyny, especially on anti-feminist online forums like Ilbe (10). Not only do these communities indicate discontent regarding females' elevated social status, but they also believe that women, especially radical feminists, exploit their growing power to gain public sympathy and undeserved privileges (29). Men upholding such ideas and thoughts are often deemed as far-right, misogynistic individuals who perceive women's new role in society as a threat to their long-established social authority within the patriarchal system (32).

According to a nationally representative sample of young Korean men, males who feel victimized possess less positive attitudes towards marriage. Interestingly, some believe marriage, the institution itself, impedes traditional male dominance due to the changing conditions and expectations of married couples with children (32). This change can be understood in the context of the "gender revolution," in which women expect their husbands to contribute equally to the household and generally share similar perceptions of gender roles (23). A study of 1,028 Korean men between the ages of 20 and 25 indicates that downwardly mobile men are more likely to express feelings of hostile sexism (33).

Another body of research demonstrates that while men claim they embrace the change in non-traditional gender roles supporting women's increased workforce participation and economic advancement, in reality, parental duties are not distributed the way contemporary South Korean females desire (25). Men's stagnant attitudes towards labor distribution, still holding that women should assume most of the household duties, are shown to influence fertility rates (25). In light of this gender gap in perspective towards household labor distribution, patriarchal societies like South Korea may experience delays in transitioning into the third 'phase' of demographic transition characterized by both equal workforce participation and gender contribution in taking care of the children. It follows that, due to men's belief that they

would be encouraged by their partners to engage in traditionally ‘motherly’ duties, potential fathers’ attitudes towards raising children have become less favorable. Such “non-pecuniary” costs, as scholars label them, could inhibit South Korean males from desiring marriage and child-rearing.

From existing arguments explaining the economic and socio-cultural conditions that influence fertility decisions, it appears that females and males hold distinct attitudes towards gender roles and family structures. Although both genders encounter similar economic hardships, such as housing and education costs, there appears to be a noticeable difference in their perspectives and experiences regarding socio-cultural factors that affect family formation choices. The following section outlines the qualitative data collection conducted for this study, which examines how the aforementioned factors contribute to South Korea’s declining fertility rate.

3. Research Methodology

To understand how various factors influence South Koreans’ fertility decisions, this study gathers qualitative data through both in-person and virtual interviews. The interview questions covered a wide range of topics and ideas, including workplace inequality, work-life balance, parental duties, societal judgment, gender conflict, generational differences, and Western influence. Nine out of ten interviewees resided in South Korea, with only one of them currently living in the United States. The participants were recruited through snowball sampling, utilizing personal and familial connections.

I interviewed four females between the ages of 31 and 41 and six males between the ages of 21 and 36. Eight interviewees belong to the upper-middle class, while two belong to the middle class. All ten participants held non-traditional attitudes towards gender roles. By non-traditional attitudes, I mean those that favor equal distribution of housework and childcare between parents and support for females’ increasing workplace participation. The interviewees’ intentions regarding having children varied; some were unsure, some were certain they wanted children, and some were certain they did not. Notably, more male participants reported wanting children than female participants. Participants represented a diverse range of occupations, including attorneys, software engineers, program managers, and a medical student. On average, each interview lasted approximately 39 minutes and 8 seconds. A table outlining important information concerning each respondent, including intent to have children, views on gender roles, and occupation, can be found below.

Interviewee	Contact Method	Gender	Age	Job	Class	Intent to have children	Gender Roles	Interview Duration (min:sec)
#1	Self	Female	40	Writer	Upper-middle	NO	Non-Traditional	44:18
#2	Self	Male	21	Seoul	Upper-middle	YES	Non-Traditi	42:02

				National University Medical Student			onal	
#3	Self	Male	27	Unemployed	Middle	UNSURE	Non-Traditional	42:19
#4	Through father	Male	36	Google Play Program Manager	Upper-middle	YES	Non-Traditional	45:14
#5	Through father	Male	33	Google Korea Lawyer	Upper-middle	UNSURE	Non-Traditional	41:09
#6	Self	Female	41	Manufacturing office worker	Upper-middle	NO	Non-Traditional	28:32
#7	Self	Female	36	Pharmaceutical company worker	Upper-middle	UNSURE	Non-Traditional	26:31
#8	Through mother	Male	35	Attorney at Shin and Kim	Upper-middle	Married and wants children	Non-Traditional	36:45
#9	Through mother	Female	31	Attorney at Shin and Kim	Upper-Middle	Married and wants children	Non-Traditional	43:03
#10	Through mother	Male	27	Software Engineer in Texas	Middle	Unmarried and wants two children	Non-Traditional	41:25

Every interview was either conducted over Google Meet or in person and was recorded by ClovaNote, an AI-powered note-taking tool that utilizes voice-to-text translation. After completing each interview, ClovaNote transcribed it in the language in which the interview was conducted, which was either Korean or English. Then, using Chat GPT premium features, I translated the interviews to English if they were not originally recorded in English. I then went through every interview transcription and translation to ensure the quotes were comprehensible and accurate. In the next section, I present the findings from the interviews, connecting them to existing claims and discussions in the preceding literature review.

4. Qualitative Findings

4.1 Economic

The interview data revealed that economic concerns play a significant role in fertility decisions, a finding consistent with the existing literature. Both male and female interviewees shared similar perspectives on the financial concerns that influenced their decision to have children, which included expensive housing and education costs, inadequate and ineffective government policies, gender wage gaps, and regional disparities, particularly between Seoul and smaller cities. The data aligns with existing research suggesting that men and women encounter similar financial barriers to fertility, despite a few differences regarding wage gaps and government policy.

4.1.1 Housing and Education

Current research suggests that among these reasons, housing costs have a particularly significant impact on deterring marital and child-rearing intentions. The housing crisis is attributed to the considerable disparity in housing prices between small and large cities in South Korea (44). Out of the OECD countries, South Korea is the only one with a housing price gap of over 200% between large and small metropolitan areas (44). Despite the greater affordability of houses in smaller cities, most families prefer to stay in large cities, where they can access enhanced education, healthcare, and job opportunities, highlighting the inseparable relationship between economic and social burdens (23). A 2022 study found that soaring housing costs remain the most significant barrier to fertility for individuals between 20 and 64 (36). In addition, another study, published in 2025, examining the relationship between homeownership, housing prices, and fertility rates, found that childbirth is delayed until individuals own a home for themselves (17). Given this, it makes sense for researchers to consider that childbirth is postponed due to soaring housing prices and the inability to achieve homeownership (17).

Unsurprisingly, eight out of ten interviewees believed unaffordable housing prices are a primary cause of low birth rates. Only two respondents, respondents 6 and 7 (both female), stated that housing policies and issues do not influence people's marital decisions or their intent to have children. Both believed that improved housing conditions will not increase fertility rates because South Koreans' mindsets have already changed to be more incompatible with the idea of forming a family. Yet, the rest of the interviewees concurred with one another in that South Korea's housing crisis has been exacerbated and left untreated by governmental policies. Some claimed that Korea lacks an effective rental system that could alleviate the financial pressures on young couples looking to have children. Interviewee 3 (male) asserted that the inflexible rental policies force citizens to "rely on either buying homes or securing long-term deposit-based leases, which are highly expensive, especially in Seoul."

Many respondents mentioned that there is enormous pressure and a significant burden to secure stable housing options in Seoul, the densely populated capital city of South Korea. Interviewee 4 (male) shared how "for [couples] in their mid-30s, it's impossible to buy a home in a neighborhood they'd want to live in—not with their own money." Two interviewees drew a comparison between housing policies and options in Daejeon, a smaller city located south of Seoul, and those in Seoul. In Daejeon, it costs nearly three times less to buy an apartment near

the city centre, which interviewee 9 (female) thought lowers the “general burden around marriage and childbirth.” Statistics from 2023 indicate that the total fertility rate (TFR) in Seoul was 0.58, while Daejeon was 0.79 (45), suggesting that lower housing costs may contribute to higher fertility rates. This raises the question of why smaller, more rural cities significantly lack the opportunities offered in Seoul and whether decreasing competition to move into Seoul will ultimately decrease housing demands and lower prices, leading to improved fertility rates.

4.1.2 Misguided Policy

Another central theme evident throughout the interviews was the notion that South Korea’s government has led misguided policies that have failed to address crucial problems concerning housing, wages, and taxes. Interviewee 3 (male) insisted that providing more financial incentives for childbirth, like some districts have executed, would motivate people to have children. Similarly, interviewee 9 (female) noted that “if there were significant benefits around housing, or if having a baby helped you afford a house, that would drastically reduce the [financial] burden.” In sum, the two respondents believed that fertility rates will increase if the government provides more support to couples and families through increased financial incentives. Moreover, interviewee 3 (male) also stated that current economic policies are targeted towards lower classes, while “anyone who earns above a certain threshold gets nothing.” They argued that such benefits must be equally given to households and couples with higher income, as they may be the ones who are “actually willing” to marry and raise children. Interviewee 4 (male) claimed that “people with...financial stability” receive little to no government support. In general, participants felt that the financial incentives offered by the government were misaligned with their needs, effectively excluding many couples who could be persuaded to have more children.

When examining recent trends and policies in Korea’s public rental housing for low-income households, studies have identified several limitations that hinder accessibility to cheonse-type housing and homeownership as a whole (30). They determined at least five limitations: questionable public rental housing support for low-income households, lack of monitoring public rental housing management after occupancy, the inflexible rent system for public rental housing, degree of fair housing assistance for various income groups, and the withdrawn attitude of local governments and private stakeholders in the public housing issue (30). Suppose the government addressed these misguided policies and structural flaws in financial support systems for young couples and potential parents. In that case, it follows that fertility rates may rise, as cheonse home ownership would become more affordable and accessible to a broader audience.

4.2 Socio-Cultural

Apart from the economic factors discussed above, including how these factors interact with culture, the interviews also touched upon socio-cultural explanations for declining fertility rates that support, modify, and contradict existing arguments. Eight recurrent themes appeared

throughout the interviews with male interviewees: hyper-competitive education, military service, patriarchal norms, gender conflicts, parenthood, comparison culture, institutional attitudes, and social media. For females, the predominant topics of discussion were distrust in society, sacrifices of motherhood, the competitive culture, unequal gender expectations, individualism, work-life balance, feminism, and the impact of social media. Thus, several overlapping themes emerged, including gender conflicts, a hyper-competitive culture, the role of the media, and the burdens of parenthood.

4.2.1 Cultural Connections to Economic Concerns

A compelling argument raised by the responses concerned the interconnections between financial issues, policies, and socio-cultural attitudes and how these factors simultaneously affected marital and fertility intentions. Interviewee 7 (male), echoing several existing studies, stood out from the rest by asserting that economic policies are unlikely to influence marital and reproductive decisions and questions the long-term effectiveness of financial incentives when cultural issues remain unaddressed (30). With regards to increased federal aid for education cost relief, some participants claim that education costs “are something you deal with after having kids,” and that it is up to individuals to decide the amount of resources they are willing to devote to children’s education. Thus, these respondents perceived that, regardless of the cost of education, those who want children will continue to do so. However, interviewees 1 and 9 (both female) disagreed, arguing that Korea’s cutthroat academic environment “is just too hard” and “makes even one child a financial burden.” They contended that South Korea has an innately intense academic culture that praises and enforces immense spending on education, making it difficult for families to bear the costs regardless of how they may personally want to raise their children. This phenomenon aligns with numerous findings that Korean parents find it difficult to avoid the financial burdens surrounding their future child’s education, as Koreans tend to judge one another by their outward degree and intensity of educational attainment (38).

Gender inequality, which includes hiring trends, in workplace environments serves as another dominant barrier to females’ reproductive decisions. According to interviewee 1(female), “hiring is unfair” in most Korean workplaces due to companies’ fear of female employees disrupting productivity if they use maternity leave upon having children. In addition, interviewee 7 (male) contended that “leadership roles are still male-dominated,” which further exacerbates wage inequality. Some positive opinions stated that eventually, women’s presence in high-paying positions will increase and reduce wage inequality. However, a larger proportion of interviewees expressed skepticism towards the dire gender wage gap in South Korea.

Together, the interview responses and studies demonstrate the futile impact of focusing solely on changing economic initiatives and government policies without tackling deeply ingrained social and cultural trends. For instance, recent data shows that South Korea’s childbirth incentives of 1 million won resulted in a barely noticeable increase in total fertility rates between 2009 and 2021 (16). Participants and studies agree that Korea’s societal perceptions and norms towards child-rearing must change in tandem with tangible policies to produce promising

improvements in fertility intentions (41). Hence, both existing studies and interview responses suggest that the government and citizens should work in collaboration to foster an environment that promotes a healthy work-life balance, enabling parents to transition freely between their work-life and household duties with minimal fear of negative consequences and workplace interruptions, especially for females. If, then, the deep-rooted culture of fierce academic competition and comparison were to shift to support the evolving perspectives of young South Koreans, there would be a greater possibility of addressing the fertility crisis.

4.2.2 Hyper-Competitive education

Concerning issues unrelated to government policies or financial matters, both females and males attributed Koreans' low child-rearing desires to the society's toxic educational arms race, a cutthroat culture fueled by academic competition, peer pressure, and unrealistic expectations. Female interviewees 1 and 9 agreed that potential mothers "worry about the stress [their] child will go through" because "kids are shuffled from one academy (cram school) to another, and it's exhausting." Interviewee 9 (female) added that if there were less pressure to invest in top-tier private education, "moms would feel less stress." Most male interviewees shared similar perspectives. Interviewee 4 (female) confessed that regardless of "how financially secure [he] might be, [he] feels [s] bad for [his] future kids," which is likely because "everyone is forced to compete within the same academic system, making education highly stressful," as interviewee three claimed. Interviewee 2 (male) offered a more objective and nuanced opinion regarding Korea's intense academic culture, citing that educational competition exists due to the opportunities for social mobility. He proposes that, compared to other developed countries, such as the United States, hard work and determination are fairly rewarded in South Korea, which increases the appeal of social mobility as a highly desirable goal. Regardless, as a result, grades, exams, especially the college-entrance exam known as the CSAT, "determine many aspects of one's economic and social standing." Moreover, Korea lacks a wide range of "successful career options," as interviewee 3 (male) believed, indicating that there are only a few prestigious job spots available for the highest-performing students. Ultimately, the participants agree that potential Korean parents opt out of having children to avoid the pressure of devoting excessive time and resources to their children's education. Therefore, Korea's hyper-competitive education culture may make childbirth undesirable.

Statistics from the Korea Herald in April 2025 confirm that Korean parents feel overwhelmed by their children's educational future, with "66.9% of parents holding themselves accountable for their children's accomplishments in life" (44). This past year, the total expenditure on private education reached its historical peak, signaling the growing tendency for households to exhaust their resources for private education, including school and supplementary classes (20). Hence, the interviewees' perspectives align with research drawing a negative relationship between ingrained socio-cultural expectations surrounding education and individual desires to have children. Furthermore, my data supports interpretations of decreased fertility intentions due to Korea's 'helicopter-parenting' culture, a form of "over-parenting" involving

parents, especially mothers, excessively controlling their children's schedule to navigate the rigorous academic day (38). To conclude, as potential parents, the interviewees echoed the many arguments made by scholars who outline a negative relationship between parental stress around children's education and the desire to have children.

4.2.3 Feminism and Gender Conflicts

Another argument mentioned throughout the interviews by both genders is that South Korea's contemporary feminism and gender conflicts hurt citizens' perceptions about marriage and child-rearing. As a whole, the interviewees characterized both modern feminism and misogyny as radical ideologies pursued by individuals who release their frustration with an issue in society or their lives by blaming other groups. Interviewee 4 (male) corroborated this claim when she stated that "misogyny.. It's an expression of discontent with [individuals'] own lives and with society." In addition, interviewee 1 (female) stressed the fact that misogynists, especially politicians, "shift attention [away from themselves] by attacking women." She believed that South Korean women, in turn, "feel disgusted and disrespected" by misogynistic comments, making them "less likely to get married or have kids." Another interviewee, interviewee 7 (female), asserted that misogynistic men "don't actually understand what feminism is" and do not "speak from true knowledge or understanding." On a similar note, interviewee 2 (male) argued that Korea's 4B movement, which opposes marriage, sex, and motherhood, represents a radical measure taken to "advocate for women's rights," which triggered reactionary responses and intensified misogynistic sentiment. In the end, it appears that the respondents generally felt that "both extremes – either hardcore feminism and the extreme opposition (misogyny) – have made marriage seem something negative." Ultimately, these extreme movements and ideologies make it more difficult for romantic partnerships, based on mutual understanding and respect, to form among contemporary South Koreans.

Taken together, the interviewees' responses regarding feminism and gender conflicts offer a new perspective that is largely unaddressed in the current literature. Current literature attributes the rise in feminism and misogyny to women's increasing socioeconomic status. However, this data suggests that these movements originate from emotional factors, such as personal frustrations with a societal issue or aspect of one's life, rather than empirically based complaints about women's advancing socio-economic status or their tendency to exploit their gender to obtain certain privileges. In conclusion, regardless of what interviewees believed causes feminism and misogyny, the responses align with several studies observing that South Korea's low fertility response rates could originate from radical sentiments rooted in these modern gender conflicts.

4.2.4 Social Media

One prevalent theme in the interviews, although not explored in the literature review, was the role of social media in deepening the severity and public perception of societal problems and gender conflicts. Respondent 10 (male) cited social media as the primary contributor to low

fertility rates because South Korea's "advanced technology and networking" and "obsession with social media" lead to extreme "ideas and thoughts, which could be false and misleading." He asserts that the rapid spread of misinformation across the country worsens online gender wars that eventually permeate into "non-internet society." Similarly, interviewee 3 (male) suggested that the "misinformation and hateful rhetoric" fuels gender conflicts and "plays a role in devaluing marriage." In his view, both genders' lack of understanding of one another and willingness to converse contribute to Korea's declining fertility desires. Interviewee 5 (male) added that dating shows and entertainment put forth the notion that "if a man gets married, his wife will take all his money, he'll work himself to death, and meanwhile, [his wife] will be relaxing... and shopping..." He concluded that Korean dramas and reality shows discourage marriage as they "paint marriage as this nightmare where men suffer."

However, on top of the media's alleged role in fostering gender-based conflicts and deterring marriage, interviewees claimed that online platforms perpetuate a culture of comparison by generating unrealistic expectations and competition. Respondents 8 and 10 (both male) claimed that parents feel elevated pressures after watching TV shows depicting "fancy [family] lives" where parents have "enough free time to go out and play with their kids." As a result, the male interviewees hypothesized that Korean couples and young individuals hesitate to have children for fear of not being capable of living up to the ideal lifestyles shown in the media.

Female respondents expressed similar concerns about the portrayal of life after marriage and childbirth in social media and public entertainment. Interviewee 9 (female) claimed that Korean mothers on social media post anxiety-provoking content, such as their kids wearing luxury clothes and attending expensive academies. Similar to the males' perspectives, female respondents admitted that the media advertises unrealistic standards and expectations. For instance, respondent 7 stated that "women are still portrayed in very traditional roles" online, where females take care of the children while men work outside. For her, the media's portrayal of wives discourages her from pursuing the traditional motherly role, which may require her to abandon marriage entirely.

Male and female respondents emphasized different aspects of the role of social media in contributing to fertility decisions. One aspect that the males omit was the role of social media in dramatizing feminism and gender conflicts. According to interviewee 6 (female), the degree of gender divide and radical feminism is depicted more intensely and negatively than in reality, causing unnecessary arguments based on rumors or exaggerated sentiments. Ultimately, the current literature aligns with these responses, citing social media – a place to showcase lavish and successful family lifestyles – as triggering anxiety, fear, and encouraging upward social comparison (39). In conclusion, the interviewees' opinions underscore the significant role that online content can play in influencing marital and fertility intentions.

4.2.5 Individualism

Another common argument made by the interviewees was that Koreans are increasingly prioritizing individual happiness and comfort over the traditional expectation of continuing the

family's lineage, which could further deter fertility desires. As interviewee 6 (female) explained, modern society puts less emphasis on family but "more about me" and fulfilling personal goals as opposed to sacrificing time and resources to getting married and raising children. This echoes published arguments indicating that more Koreans view marriage as optional, not mandatory, unlike in past generations, where families strongly enforced marriage and child-rearing (35). Interviewee 8 (male) elaborated on this idea, citing how Korea's traditional Confucian values, centered around the family, have not yet adapted to modern individualistic mindsets. Hence, interviewees 6, 7, and 8 concluded that young Koreans feel less obliged and willing to have children because they refuse to give up opportunities for self-development. Indeed, interviewee 8 further developed this claim by reasoning that the combination "of individualistic attitudes...peer pressure and the competitive society" produces a discouraging environment for marriage and child-rearing, as citizens doubt their ability to juggle career demands, family life, and personal leisure at once.

Furthermore, the ten interviewees generally held non-traditional attitudes towards gender roles. Traditional beliefs support the concept of husbands as 'breadwinners' who bolster the family's financial stability, while wives raise the children and perform household chores. Most interviewees claimed that mothers and fathers should share equal responsibilities both within and outside the household. Several respondents insisted that "both partners [should] do everything they can" and "give their best" in parental duties. Thus, the interviewees' opinions reaffirm existing research, which posits that Confucianist norms go against contemporary gender-role perspectives, hindering fertility and marital intentions. However, the males' responses suggest that young Korean men are not all conservative and express disapproval towards women's desire for equal housework distribution; in this study, all male respondents hoped to contribute equally to household duties alongside their future or current wives. Thus, the interview responses contradict some scholars' arguments that claim that Korean men's tendency to support conservative gender roles augments females' reluctance towards having children and limits their presence in the labor market (25). While the sample is small and limited in diversity, the responses nevertheless introduce a possibility that it is not gender-based ideological conflicts that deter marriage and child-rearing, but relatively stagnant societal values accompanied by rapid individual mindset shifts that seem to taint the prospect of having children.

4.2.6 Overworking culture

One of the final arguments that emerged throughout both male and female interviews was South Korea's overworking culture and its adverse effects on the perception of parenthood and family life. Interviewees believed that the lack of work-life balance, resulting from the implicit pressures to work overtime, reduces the time available to spend with partners and family members. Well-aware of this trend, potential parents doubt their ability to simultaneously manage their work and household responsibilities, which, according to interviewee 9 (female), "raises the barrier to having children" and adds "psychological pressure around marriage and childbirth." However, numerous interviewees mentioned that their friends working in "public

service or large corporations often have longer parental leave and flexible hours for childcare,” making parenthood seem less stressful and burdensome. On the other hand, interviewees currently working for the private sector, such as law firms, stated that their “working hours are long and irregular.” Interviewee 8 (male), who works for a prestigious law firm in Seoul, expressed that “work-life balance in any law firm in Korea” is a “mythical thing that everybody speaks, but it’s a holy grail that [they] don’t ever get.” In the future, it would be valuable to determine if there is a gap in fertility rates between those employed in lower-paying, public-sector jobs and their counterparts in private workplaces.

In addition to the differences between the workplace cultures in public and private corporations, a contrasting set of social norms appears to exist between American companies and Korean companies in Korea. For instance, interviewee 7 (male), who works for an American company composed mostly of females, noted that her company has “flexible hours so people with families can work around their schedules.” Then, she explained that in “most Korean companies, people do feel pressure when taking [parental] leave” because employees know that it is challenging to find replacements, unlike in American companies, where bosses can find international replacements to cover for domestic workers. Hence, according to the interviews, American companies in South Korea provide a more friendly environment for potential parents who want children. As interviewee 9 (female), who works at another leading law firm in Seoul, suggested, companies must implement effective leave policies and accommodations, especially for females, such as “strong policies saying that women cannot face disadvantages (career interruptions like lower wages) when returning from leave.”

Nevertheless, research indicates that parental leave-related policies do not necessarily alleviate pressure on employees who are considering starting a family (25). Due to the prevailing hard working culture, which affects both genders, but particularly females, individuals may still be reluctant to take time away from work to pursue their personal lives (25). Following existing claims, the interview arguments suggest that the entrenched culture of hard work and workplace obedience hinders the effectiveness and real-life implementation of various policies. As a result, individuals may resort to abandoning marriage and parenthood altogether to showcase their industriousness and willingness to advance in the workplace hierarchy. In this context, scholars acknowledge that raising children limits opportunities to be viewed favorably and receive fair opportunities for advancing in one’s occupational position.

4.2.7 Culture of Comparison and Peer Pressure

The final prominent topic amongst the interviews was the role of the pervasive culture of comparison and peer pressure in discouraging marriage and family formation. Nearly all interviewees expressed disapproving attitudes towards these cultural phenomena, with some acknowledging that the comparison culture and peer pressure contribute to lower happiness levels and pessimistic outlooks on life. Interviewee 10 (male) attributed Korea’s culture of comparison to the nation’s long-established and generally praised “nunchi” culture, which he believes plays an immense role in deterring marriage and child-rearing intentions. To possess

“nunchi” typically involves observing, analyzing, understanding, and responding to subtle cues to act appropriately in a given social setting. Interviewees 3 and 8 (both male) mentioned that practicing skilled nunchi necessitates paying an unhealthy amount of attention to others’ judgments and perception. Moreover, having “nunchi” could mean submitting to external standards and expectations. These unrealistic and harsh expectations, as asserted by interviewees 1 (female), 2 (male), 3 (male), and 4 (male), include attending only the top universities in Seoul, living in luxurious apartments, providing an ideal educational environment for their children, and pursuing high-paying, prestigious occupations. Enforced to be conscientious and adhere to prevailing norms, a growing number of Koreans remain unmarried or childless to avoid the burden of fitting in. The interviewees’ responses support the existing literature, which argues that Korean society has rigid socio-cultural norms that persistently dictate individual thoughts and behaviors. Moreover, failing to meet standards frequently results in both implicit and explicit judgment and criticism (39). Thus, it is unsurprising, after reviewing existing claims and conducting the interviews, that the unwavering culture of interpersonal comparisons decreases contemporary individuals’ child-rearing desires.

4.3 Males

Apart from the numerous commonly mentioned arguments by males and females, the interviews yielded a disparate set of ideas and arguments highlighted exclusively by male respondents. Some of the dominant topics of discussion included the role of Korea’s mandatory military service and patriarchal norms on marital and child-rearing intentions for South Korean men.

4.3.1 Military service

The concept of ‘victimhood’ was mentioned several times as a possible deterrent to child-rearing intentions. One reason Korean males might feel ‘victimized’ and mistreated is due to the mandatory military service required solely for males. Interviewees 3 and 5 believed that Korea’s current military service system lacks practical opportunities and vocational training for young men to transition smoothly back into the workplace after completing their military service. Interviewee 3 predicted that improved employment policies for men in their post-military years could make marriage and having children less overwhelming, since the men may be financially stable with a full-time job. Hence, male interviewees believed that the poor system surrounding the military service, and the fact that it is required only for male citizens, increase sentiments of ‘victimization.’ Some existing literature finds that male victimhood among Korean men is negatively correlated with marital intentions (32). Men’s feelings of frustration, betrayal, and exclusion may stem from Korea’s unique military service system and its inadequate provisions for supporting young males as they move into post-service life in the workplace and society. However, it is critical to note that the concept of male victimhood was not cited as a significant factor affecting fertility decisions in the interviews.

4.3.2 Patriarchal Norms

Apart from male victimhood, another contributing factor for males' hesitance to have children was burdensome paternal expectations and norms. One of the collective concerns was the inability to freely use paternal leave in the workplace to support their wives and households. As interviewee 4 asserted, males feel the need to "earn more, provide more..." and therefore believed males cannot use paternity leave without feeling guilty or facing criticism from colleagues and bosses. Although they acknowledged that Korea has become more supportive of non-traditional parental role distribution, most companies still expect men to work industriously all the time, resulting in a lower proportion of males taking parental leave compared to females (46). Interviewee 5, who formerly worked at a leading Korean law firm, claimed that "the number of male employees who've taken parental leave is still zero." His comment suggests that South Korea, despite individuals increasingly embracing non-traditional attitudes towards work-life balance and gender roles, still possesses a patriarchal society where "men are expected to be good [financial] providers."

However, on the other hand, in American companies, "taking parental leave isn't something to feel guilty about..." according to interviewee 5, who transitioned from a leading Korean firm to a leading American firm. While current research lacks information on how parental leave is viewed differently in American and Korean companies, the interview responses affirm that men often feel that going on leave would hinder their career advancement (39). In conclusion, both current literature and interviews highlight that Korean companies' persisting patriarchal and harsh work culture leads to unequal parental leave use, impeding the opportunity for men to adopt a balanced lifestyle that allows them to contribute more to their household alongside their wives.

4.4 Females

Similar to male interviewees, female interviewees felt they are faced with unique barriers and factors affecting their reproductive decisions. Their main argument was that individuals' distrust in society and norms surrounding motherhood and gender roles have the most significant effect on fertility intentions.

4.4.1 Distrust in society

While it is unlikely that distrust in society is exclusively felt by females, in this study, the female interviewees were more likely to express pessimistic opinions about South Korea's overall society and future, as well as the consequences of their distrust on fertility intentions. Interviewee 1 frequently noted that "women are losing trust in society and in the idea of a fair community..." which then discourages females from wanting to raise children in Korean society that they perceive as particularly detrimental for women. She explained that Korea's "overly competitive and unstable" society reduces overall happiness and weakens child-rearing intentions. These opinions corroborate scholarly claims that young adults' uncertainty towards

their society's future economy and equality may undermine their future child-rearing intentions (39).

4.4.2 Motherhood and Unequal Gender Expectations

Expectations surrounding motherhood, unlike distrust in Korean society, appears to be a factor influencing many interviewees' perceptions towards having children. Respondents perceived that once women come back from maternity leave, they will face subtle, if not overt, instances of discrimination and unfair treatment by their companies, such as receiving less-paying work and experiencing stalled promotion. Additionally, in line with several published claims, several participants argue that wage inequality could originate from women's voluntary absence in demanding jobs to devote greater time towards housework and child-rearing (48). They explained that women still feel obliged to take care of the children themselves and handle the bulk of chores, which inclines them towards seeking flexible jobs with friendlier working hours and culture – a phenomenon referred to as motherhood penalty.

Korea's motherhood penalty reflects how South Korean institutions and society, as a whole, still consider women as primary caretakers, even while overtly supporting females' robust workforce participation. Extant literature confirms that perceived career interruptions, such as implicit and explicit workplace discrimination, additional pressure on working mothers, and unfair workplace cultures discourage fertility desires (10). Therefore, apart from more effective government policies, socio-cultural changes seem necessary to alleviate fears of motherhood penalties, including career disadvantages and poor work-life balance. Despite more females and males holding non-traditional attitudes regarding Korea's work culture and the gender distribution of household responsibilities, the national culture appears to dominate over evolving individuals' beliefs.

5. Conclusion

This study has examined the overarching question: Why are South Koreans choosing not to have children? The answers are categorized into economic and socio-cultural reasons and further differentiated by gender. The study revealed that the main financial factors negatively affecting fertility decisions are expensive housing costs and education expenses, which current policies fail to address. These reasons corroborated extant literature, which states that South Koreans remain unmarried or childless due to fears of not being able to afford housing in Seoul or to provide their children with high-quality education. Regarding socio-cultural aspects, this study revealed that South Korea's educational arms race, hypercompetitive society, reliance on social media, patriarchal culture, and gender inequality all deter fertility intentions. Such findings once again were consistent with existing arguments that attribute South Koreans' reluctance towards child-rearing to the prevailing systemic pressures and socio-cultural expectations that increase childlessness.

Most importantly, however, this study highlights the interdependence between economic and socio-cultural factors, suggesting that improving economic issues (i.e.; soaring housing

costs) would first necessitate changes in cultural aspects, including the competitive academic environment, patriarchy, and tendency to engage in interpersonal comparisons. Furthermore, this study posits that a significant gap exists between stagnant socio-cultural structures and beliefs – such as viewing women as caretakers, men as breadwinners, valuing prestige and name value – with more contemporary desires to lead a balanced lifestyle both at work and at home. This finding supports several researchers' opinions that the existence of this gap may explain the growing reluctance to raise children.

However, the study also contradicts a body of research that establishes a linear relationship between gender wars, the increase of radical, conservative young males, male victimhood, and declining fertility rates. The interview responses present that these factors only play a modest, if not insignificant, role in influencing marital and child-rearing intentions. Hence, this study proposes prioritizing gradual reform in Koreans' competitive mindsets, the society's hyper-sensitive environment, and deep-rooted patriarchal system.

Regardless of the findings, it is imperative to consider this study was based on ten qualitative interviews with South Koreans from privileged backgrounds, most of whom held non-traditional attitudes towards gender roles. Further studies would benefit from interviewing and studying the opinions of South Koreans belonging to a variety of ages, social classes, and ideological backgrounds. Moreover, future research could examine young adults' attitudes specifically, exploring the paths they take to develop particular perspectives on having children in South Korea. Another study might involve interviewing employees in different regions, hired by disparate companies, and engaged in distinct work sectors, to determine how fertility beliefs depend on an individual's environment and occupation. To fully comprehend the exhaustive narrative surrounding South Korea's dire birth rate crisis, scholars should continue gathering more qualitative and quantitative data. An increasing number of studies and analyses may produce optimal ways to restructure deep-rooted cultures and norms without excessively burdening institutions, people, and the government.

Works Cited

- Alderotti, Giammarco. 2022. "Female Employment and First Childbirth in Italy: What News?" *Genus* 78 (10). <https://doi.org/10.1186/s41118-022-00162-w>.
- Anderson, Thomas, and Hans-Peter Kohler. 2013. "EDUCATION FEVER and the EAST ASIAN FERTILITY PUZZLE." *Asian Population Studies* 9 (30): 196–215. <https://doi.org/10.1080/17441730.2013.797293>.
- Arpino, Bruno, Gøsta Esping-Andersen, and Léa Pessin. 2015. "How Do Changes in Gender Role Attitudes towards Female Employment Influence Fertility? A Macro-Level Analysis." *European Sociological Review* 31 (16): 370–82. <https://doi.org/10.1093/esr/jcv002>.
- Barro, Robert J., and Jong Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104 (15902): 184–98.
- Bi, Xiyan. 2024. "Research on the Factors Contributing to Low Fertility in South Korea." *Lecture Notes in Education Psychology and Public Media* 62 (10): 34–38. <https://doi.org/10.54254/2753-7048/62/20241675>.
- Briselli, Giulia, and Libertad Gonzalez. 2023. "Are Men's Attitudes Holding Back Fertility and Women's Careers? Evidence from Europe." https://conference.nber.org/conf_papers/f178420.pdf.
- Capelos, Tereza, Ellen Nield, and Mikko Salmela. 2023. "Narratives of Success and Failure in Ressentiment: Assuming Victimhood and Transmuting Frustration among Young Korean Men." *Social Sciences* 12 (32): 259. <https://doi.org/10.3390/socsci12050259>.
- Chi, On. 2025. "Can Government Financial Aid Save an Aging Society? A Systematic Review on the Effect of Financial Incentives on the Fertility Rate." *Open Journal of Business and Management* 13 (02): 1018–46. <https://doi.org/10.4236/ojbm.2025.132055>.
- Cho, Duksang, and Jungmin Han. 2024. "Women's Career Interruptions and the Declining Fertility Rate in South Korea - KDI - Korea Development Institute - RESEARCH - KDI FOCUS." *Kdi.re.kr*. 2024. https://www.kdi.re.kr/eng/research/focusView?pub_no=18386.
- Cho, Kyung Ae. 2021. "Korea's Low Birth Rate Issue and Policy Directions." *Korean Journal of Women Health Nursing* 27 (10): 6–9. <https://doi.org/10.4069/kjwhn.2021.02.16>.
- Choi Jeong-yoon. 2025. "Private Education Spending Hits Record despite Fall in Student Numbers - the Korea Herald." *The Korea Herald*. March 13, 2025. <https://www.koreaherald.com/article/10440947>.
- Choi, Jasmine. 2024. "Survey Shows Shifts in South Korean Attitudes towards Marriage and Family." *Businesskorea*. November 12, 2024. <https://www.businesskorea.co.kr/news/articleView.html?idxno=229182>.
- Chung, Hee Soo, and Jeong Ho Kim. 2004. "Housing Speculation and Housing Price Bubble in Korea." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.535882>.
- Daily, The Chosun. 2025. "Housing Crisis in S. Korea: A Growing Divide between Large and Small Cities." *The Chosun Daily*. 2025.

- <https://www.chosun.com/english/national-en/2025/02/21/C3Q3MWKA3BARLJJ4VJC3QMU3AI/>.
- Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt. 2022. "The Economics of Fertility: A New Era." SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4086239>.
- Kim, Joeun. (2023). The Gender War and the Rise of Anti-family Sentiments in South Korea. 10.1007/978-3-031-29666-6_9.
- Medwid, Jessica Daisy, "The East Asian Fertility Crisis: Causes, Consequences, and Policy Responses" (2024). ALL - Honors Theses. 38.https://scholarsarchive.library.albany.edu/all_honors/38
- Xu, B., Pak, M. Child-raising cost and fertility from a contest perspective. *Public Choice* 186, 9–28 (2021). <https://doi-org.ezproxy.oberlin.edu/10.1007/s11127-019-00751-y>
- 동아일보. 2025. "Korea's Birth Rates Make the First Turnaround in Nine Years | the DONG-A ILBO." 동아일보. January 30, 2025. <https://www.donga.com/en/article/all/20250131/5424493/1>.
- "Equality Questions: Jobs for the Boys under Fire in South Korea." 2018. RTL Today. November 14, 2018. <https://today.rtl.lu/news/business-and-tech/a/1266798.html>.
- Esteve, Albert, Diederik Boertien, Ryohei Mogi, and Mariona Lozano. 2020. "Moving out of the Parental Home and Partnership Formation as Social Determinants of Low Fertility." *Vienna Yearbook of Population Research* ... 18 (July). <https://doi.org/10.1553/populationyearbook2020.deb02>.
- Exploring Norway's Fertility, Work, and Family Policy Trends. 2023. OECD. <https://doi.org/10.1787/f0c7bddf-en>.
- Hae Won Kim, and Seo Yun Kim. 2023. "Gender Differences in Willingness for Childbirth, Fertility Knowledge, and Value of Motherhood or Fatherhood and Their Associations among College Students in South Korea, 2021." *Archives of Public Health* 81 (10). <https://doi.org/10.1186/s13690-023-01127-x>.
- He, Ruoshui, and Sofia Chen. 2022. "Socio-Economic and Demographic Factors Associated with Fertility -Southeast and East Asian Evidence." <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9082127&fileId=9082128>.
- Hwang, Jisoo. 2023. "Later, Fewer, None? Recent Trends in Cohort Fertility in South Korea." *Demography* 60 (30). <https://doi.org/10.1215/00703370-10585316>.
- Ivanova, Katya, and Nicoletta Balbo. 2024. "Societal Pessimism and the Transition to Parenthood: A Future Too Bleak to Have Children?" *Population and Development Review*, March. <https://doi.org/10.1111/padr.12620>.
- Jin, Mee-Youn, and Seok-Je Lee. 2013. "Korean Public Rental Housing for Low-Income Households: Main Outcome and Limitations." *LHI Journal of Land, Housing, and Urban Affairs* 4 (34): 303–16. <https://doi.org/10.5804/lhij.2013.4.4.303>.

- Joon-hyun, Moon. 2025. "Outdated Gender Roles Keep South Korea's Wage Gap Stubbornly High: Labor Union Survey - the Korea Herald." *The Korea Herald*. March 9, 2025. <https://www.koreaherald.com/article/10437088>.
- Jung, Han Wool. 2023. "A New Variation of Modern Prejudice: Young Korean Men's Anti-Feminism and Male-Victim Ideology." *Frontiers in Psychology* 14 (23). <https://doi.org/10.3389/fpsyg.2023.1230577>.
- Kim, Eungyung, and Jee-Seon Yi. 2024. "Determinants of Fertility Intentions among South Koreans: Systematic Review and Meta-Analysis." *Behavioral Sciences* 14 (29): 939–39. <https://doi.org/10.3390/bs14100939>.
- Kim, Ji-yun, and Miseon Jang. 2025. "Evaluating Housing Policy Effects on Childbirth Intentions in South Korea: Preferences, Benefits, and Policy Implications." *Population Space and Place* 31 (34). <https://doi.org/10.1002/psp.70036>.
- Kim, Joeun. 2023. "Antifeminist Sentiments and Marital Desire among Young Men: Evidence from South Korea." *Socius* 9 (January). <https://doi.org/10.1177/23780231231196791>.
- . 2024. "Men's Decline and Rising Support for Hostile Sexism: A Survey Experiment from South Korea." *Social Science Research* 127 (December): 103134. <https://doi.org/10.1016/j.ssresearch.2024.103134>.
- Lee, Jihoon, and Hyun Young Yi. 2025. "South Korea's Policy Push Springs to Life as World's Lowest Birthrate Rises." *Reuters*, February 26, 2025. <https://www.reuters.com/world/asia-pacific/south-koreas-policy-push-springs-life-worlds-lowest-birthrate-rises-2025-02-26/>.
- Lee, Misun, Zuzanna Brzozowska, and Tomáš Sobotka. 2024. "Declining Appeal of Marriage? Trends in Marital Intentions in South Korea, 1991–2018." *Asian Population Studies*, December, 1–23. <https://doi.org/10.1080/17441730.2024.2428026>.
- Main Findings from the 2022 OECD Risks That Matter Survey. 2023. OECD. <https://doi.org/10.1787/70aea928-en>.
- Myong, Sunha, JungJae Park, and Junjian Yi. 2020. "Social Norms and Fertility." *Journal of the European Economic Association* 19 (32): 2429–66. <https://doi.org/10.1093/jeea/jvaa048>.
- OECD. 2025. "Fertility Trends across the OECD: Underlying Drivers and the Role for Policy." OECD. 2025. https://www.oecd.org/en/publications/society-at-a-glance-2024_918d8db3-en/full-report/fertility-trends-across-the-oecd-underlying-drivers-and-the-role-for-policy_770679b8.html.
- Park, Soojin Oh. 2025. "Why the World's Lowest Birth Rate Prevails in Korea and Defy Billions of Public Investments: A Critical Discourse Analysis of Parenting, Child Care, and Family Life," April. <https://doi.org/10.20944/preprints202504.1579.v1>.
- Rashid, Raphael. 2023. "South Korea Has so Few Babies It Is Offering New Parents \$10,500." *Www.aljazeera.com*. April 12, 2023. <https://www.aljazeera.com/news/2023/4/12/south-korea-splashes-the-cash-in-scramble-to-fix-fertility-crisis>.

- Roser, Mas. 2014. "The Global Decline of the Fertility Rate." Our World in Data. 2014.
<https://ourworldindata.org/global-decline-fertility-rate#article-citation>.
- Shepherd, Tory. 2024. "Birthrates Are Plummeting Worldwide. Can Governments Turn the Tide?" The Guardian. The Guardian. August 10, 2024.
<https://www.theguardian.com/world/article/2024/aug/11/global-birthrates-dropping>.
- "The Digest Economic Growth, Cultural Traditions, and Declining Fertility." 2025.
https://www.nber.org/sites/default/files/2025-03/apr25_0.pdf.
- Yoon Min-sik. 2025. "'2 out of 3 Korean Parents Feel Responsible for Child's Success or Failure' - the Korea Herald." The Korea Herald. April 9, 2025.
<https://www.koreaherald.com/article/10461264>.
- Yoon, L. 2023. "South Korea: Fertility Rate in Daejeon 2023 | Statista." Statista. 2023.
<https://www.statista.com/statistics/1336368/south-korea-fertility-rate-in-daejeon/>.
- Yum, Minchul, and Daisoon Kim. 2025. "The Effects of Parental Leave Policy Reforms on Fertility and Gender Gaps." CEPR. March 17, 2025.
<https://cepr.org/voxeu/columns/effects-parental-leave-policy-reforms-fertility-and-gender-gaps>.

Diving into Innovation: Utilizing Kingfisher and Gannet Adaptations for Rocket Nose Cone Design By Lauren Johnson

Abstract

This study explores the potential of biomimicry in aerospace engineering by analyzing the aerodynamic benefits of rocket nose cones inspired by the natural diving mechanisms of kingfishers and gannets. These birds exhibit exceptional abilities to transition smoothly between air and water at high speeds, offering promising models for reducing drag and turbulence in rocket flight. Using 3D CAD modeling based on bird skull morphology and computational fluid dynamics (CFD) simulations in SimScale, this research compares the aerodynamic performance of the biomimetic designs against traditional nose cone shapes employed by NASA and SpaceX. Results show that the kingfisher- and gannet-inspired nose cones significantly reduced drag and turbulent kinetic energy, rivaling or surpassing conventional designs such as the ogive nose cone. The kingfisher model, in particular, exhibited the lowest turbulence across all speeds tested, while the gannet model demonstrated the second-lowest drag coefficient. These findings highlight the effectiveness of nature-inspired engineering solutions and suggest that incorporating biological principles can enhance rocket efficiency, stability, and reusability. The study underscores the need for further experimental validation through physical testing and expanded analysis of additional flight variables such as heat resistance and structural integrity.

Introduction

With a flash of iridescent feathers, the Kingfisher plunges from 30 ft; its body slices through the air like a dart. At speeds of up to 25 mph, the bird dives from the sky with such precision and grace that it enters the water with barely a ripple. The Kingfisher's sharp, tapered beak and streamlined body allow it to seamlessly transition from air to water. In this elegant dive, the bird's design maximizes efficiency, a natural wonder that speaks volumes about the evolution of form and function in the animal kingdom.



Figure 1: Kingfisher (Fisher, 2016)

Similarly, the Gannet - a seabird nearly as large as an albatross - soars hundreds of feet above the ocean, its wings spread wide against the open sky. Then, in an instant, it tucks its wings tightly to its body and plunges toward the water at a staggering 60 mph. It crashes into the water with an unbelievable force. Yet, the bird emerges unscathed, its sleek body perfectly designed to absorb the tremendous impact. With a quick shake of its wings, it takes to the sky once more to repeat the remarkable feat.



Figure 2: Gannet (The Gannet's Dive | Was It Designed?, 2025)

These animals' astounding ability to seamlessly cut through air and water offers an exceptional model for human engineering.

Biomimetics, or biomimicry, uses nature-inspired design to solve complex problems. Examples of this can be found everywhere. The airfoil shape of plane wings was taken from the design of birds, Velcro was created to mimic prickly plant burrs that stick to clothes and skin, and even solar power was engineered to replicate photosynthesizing plants. (Biomimicry: Natural Designs - Activity, n.d.) Engineers use biomimicry because nature is already optimized through thousands of years of evolution; engineers only need to apply it to a new problem. Biomimicry from Kingfishers and Gannets can be applied to rockets' challenges: drag, instability, and pressure distribution.

In particular, there is an ever-growing demand for cheaper and more efficient space travel with communication, navigation, weather, agricultural, space exploration, scientific discovery, and more relying on rockets. The private space company, SpaceX, has revolutionized the space industry with its reusable rockets, cutting costs to \$1,200 per pound or \$67 million per launch. (Ward, 2024) However, although the cost has decreased, the price tag of sending objects into space is still too inconvenient. Therefore, modeling nose cones from nature could allow for more aerodynamic designs, reducing fuel and cost. Additionally, these designs would be specially optimized for travel to planets or moons with dense atmospheres, such as Venus or Titan, that

require a nosecone to displace air particles and reduce the increased heat caused by friction. As seen when a Kingfisher dives from a height, its pointed beak pierces the water's surface with a minimal splash, while its compact shape reduces drag and enables a smooth entry. This unique adaptation could be a potential model for designing rocket nose cones. The elongated, pointed shape of the kingfisher's beak helps distribute pressure evenly during impact, reducing resistance in both air and water. A rocket's nose cone modeled after the kingfisher could offer significant aerodynamic benefits, reducing turbulence and drag during flight while enhancing structural integrity during water landings and reducing cost. Likewise, Gannets' long, narrow body shape minimizes drag during high-speed dives and protects their internal structure from the force of hitting water at such velocities. The design element of the Gannet's shape could be applied to rocket bodies and nose cones to help dissipate force upon reentry and landing, making it particularly relevant to improve rocket safety and reusability. Therefore, emulating the Kingfisher's and Gannet's shape and dive, engineers can design rocket components that handle abrupt changes in the environment—from the vacuum of space to Earth's dense atmosphere—increasing speed and withstanding structural damage. This leads to the central research question: Will a rocket nose cone based on the biomimicry from Kingfishers and Gannets reduce drag and instability?

Literature Review

Current models

Modern rocket nose cone designs face two key aerodynamic challenges: drag, which increases fuel consumption as air resistance opposes the motion of the vehicle, and instability, which causes rough transitions through varying atmospheric pressures and turbulence by unevenly distributing forces (Hall, 2023; Sia, 2021). Several existing designs try to negate these challenges although they have several drawbacks. “For example, the parabolic/rounded nose is optimal for subsonic speeds (below Mach 0.8) as it minimizes drag to increase altitude. Unfortunately, this configuration performs less well at supersonic speeds due to [the formation of shock waves]” (Basics of Rocket Aerodynamics Optimization, n.d.), according to The Peak of Flight newsletter, published by Apogee Components, a well-known company in the model rocketry. They further articulate the simplicity and cost-effectiveness of the conical nose cone, although it is not optimized to perform at transonic and supersonic speeds. Additionally, the more slender Ogive nose cone performs better at transonic and supersonic speeds, however, it is more difficult to manufacture. Lastly, they describe the Von Kármán Nose Cone, which decreases drag at transonic speeds but can be unstable and turbulent. (Basics of Rocket Aerodynamics Optimization, n.d.)

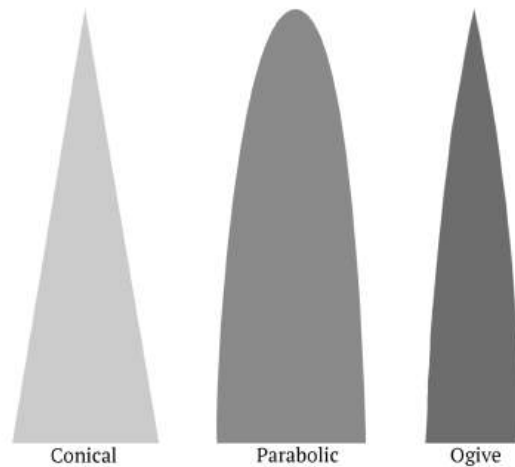


Figure 3: Nose Cones

These current designs demonstrate the need for a nose cone that can reduce drag and instability. The article *Rocket Physics, the Hard Way: Re-entry and Hypersonic Flight* by Jess Sia argues, “Understanding shock physics and geometry is crucial to aircraft and rocket design. At the speeds at which rockets fly through the atmosphere, shocks can cause parts of the rocket to overheat from the high temperatures or to suffer structural failure from the increased pressures. They also increase the drag generated by the vehicle, meaning that the engines have to work harder to push the rocket through the atmosphere.” Therefore, by mimicking the streamlined beak shapes of Kingfishers and Gannets, rocket nose cones can be designed to minimize shockwave intensity, reducing drag, overheating, and structural stress during high-speed atmospheric flight. Modern advancements in rocket nose cone design continue to address the challenges of drag and aerodynamic stability. While existing nose cone shapes offer trade-offs between efficiency, manufacturability, and performance at different speeds, there is still significant room for innovation. The need for a shape that can effectively minimize shockwave formation while maintaining structural integrity and reducing fuel consumption remains a crucial focus in aerospace engineering. By turning to nature for inspiration, biomimetic designs—such as those modeled after the streamlined beaks of Kingfishers and Gannets—present a promising solution. These birds have evolved to pierce through air and water with minimal resistance, a characteristic that could be adapted for rocket nose cones to optimize performance across subsonic, transonic, and supersonic speeds. As research continues into shockwave mitigation and material advancements, the integration of biomimicry with modern aerodynamics has the potential to revolutionize spacecraft efficiency and safety. By reducing energy loss due to drag and minimizing heat accumulation, future rocket nose cones could contribute to lighter, more fuel-efficient, and resilient spacecraft, ultimately making space travel more sustainable and accessible.

Biomimicry

Several researchers have discussed ways to combat the issues in aerospace design, one of which is using biomimicry to provide new solutions that traditional engineering approaches might not consider. “Humans' initial desire for flight stems from the imitation of flying creatures (e.g. insects, bats, and birds) in nature. Research on the aircraft that simulates flying biological organisms has been reported for centuries” acknowledges Jiakun Han et al. from the University of Xi'an, China. (HAN et al., 2020) They demonstrate how nature has become the crucial foundation of flight systems, and was even used by Leonardo Da Vinci in the Renaissance as he designed the inspiration for the very first airplanes. In addition, researchers from Chung-Ang University indicate the applications of engineering with biomimetics include designing energy-efficient vehicles inspired by the aerodynamics of birds, creating durable materials modeled after spider silk, and developing advanced robotics that mimics the movement and adaptability of animals. (Choi et al., 2015) They further explain the effectiveness of applying nature-inspired approaches to aerospace as well as other engineering fields. Similarly, Dr. Mark Ayre from the Advanced Concepts Team at the European Space Technology Centre, in Noordwijk, Netherlands emphasizes “The use of artificial muscles (e.g. electro-active polymers, shape memory alloys etc.), biomimetic image processing and vision systems, biologically inspired guidance and control for flight and ambulation, drag reduction mechanisms, cooperative guidance algorithms for intelligent swarms, flapping wing flight, smart aerospace structures, and so forth.” (Ayre, n.d.) His research establishes the idea of an intelligent biomimetic nose cone, using artificial muscles to alter the shape of the rocket to optimize it for different stages of flight. The notion of Dr. Ayre is supported by the research of Jieliang Zhao et al. at the Department of Mechanical Engineering, Tsinghua University, Beijing. In their research, they explore a morphing biomimetic nose cone design based on the abdomen of a honeybee, designed to bend and stretch during flight. They concluded, “The morphing nose cone could complete telescoping movement and bending deformation, and further took effect on reducing the drag in the reentry stage and lowering the aerodynamic heat in the high-speed gliding stage.” Although this design could reduce drag and heat, it remained unstable because of the flexible parts. Despite its instability, the morphing nose cone design inspired by the honeybee abdomen provides valuable insights into how biomimicry can revolutionize aerospace engineering. Reducing drag and aerodynamic heat during flight opens the door for further exploration into more stable and efficient designs that still reduce heat and drag, such as nose cones modeled on the streamlined and rigid structures of Gannets and Kingfishers, which are naturally optimized for high-speed dives and precision.

Kingfishers & Gannets

Kingfishers and Gannets are ideal animals to model a rocket nose cone after due to their exceptional aerodynamic adaptations, enabling them to pierce through air and water with minimal drag, stability, and pressure distribution. This remarkable ability is supported by research highlighting the unique morphology of these birds, which minimizes drag and impact forces during high-speed dives. Crandell et al. from the School of Natural Sciences at Bangor

University determined, “Bird beaks appear well-adapted to avoid both aerodynamic and hydrodynamic drag.[...]Accelerometers mounted to free-living Cape Gannets sampling at 16–32 Hz detected no or minimal deceleration due to impact during foraging dives. Drag reduction due to morphology may help reduce immediate impact forces.” These findings show that the shape of the Gannet's beak, neck, and skull allows a smooth transition between the air and water with minimal drag and deceleration. This can be applied to a rocket nosecone to allow air to flow seamlessly around the shape, reducing drag as the rocket enters different atmospheric densities in flight that could slow it down. Crandell's ideas are supported by the Japanese Shinkansen bullet trains, connecting numerous cities in Japan. “Despite being able to travel at speeds of up to 200 miles per hour (320 kph), they have also proved incredibly safe over their 50-year-plus history. But the early trains had one major problem: they created sonic booms when they left tunnels.” (Engineering, n.d.) To solve this problem, the engineers took inspiration from the Kingfisher's streamlined body that can pierce through the air and water, making little to no splash as it catches its prey. The new Kingfisher-inspired design allowed the train to travel fast while redistributing air to reduce the impact of the sonic booms. This application to trains can also be applied to rockets as a sonic boom can impact the rocket's stability when they accelerate to high speeds. With a nose cone modeled after a Kingfisher, the pressure will be distributed similarly to the Shinkansen bullet train, allowing the rocket to maintain balance. By drawing inspiration from the aerodynamic adaptations of Kingfishers and gannets, engineers can design rocket nose cones that minimize drag, reduce impact forces, and improve stability during flight. Just as these birds seamlessly transition between air and water and the Shinkansen train mitigates sonic booms, a biomimetic nose cone can enhance a rocket's performance by optimizing pressure distribution and reducing deceleration through varying atmospheric densities. This natural design approach offers a promising solution for improving the efficiency and stability of high-speed aerospace vehicles.

Overall, the lack of research focusing on applying biomimetic designs, specifically those of Kingfishers and Gannets, to aerospace vehicles prompts the question of whether a rocket nose cone based on the biomimicry of kingfishers and gannets reduces drag and instability.

Methodology

Furthermore, to answer the question, will a rocket nose cone based on the biomimicry from kingfishers and gannets reduce drag, thermal stress, and instability?” I created a quantitative experimental method study. I measured the drag and stability variables in a computational fluid dynamics (CFD) simulator. These variables were carefully selected because NASA has identified drag and stability as critical forces that significantly impact a rocket's performance during flight. Additionally, these factors directly correlate with the natural adaptations of Kingfishers and Gannets, which have evolved to minimize drag, manage impact forces, and maintain stability during high-speed dives. By analyzing these variables, the study aimed to determine whether biomimicry from these birds can effectively mitigate the challenges rockets face when transitioning through different atmospheric conditions. (Hall, 2023; Kingfisher

Diving - How Fast Does a Kingfisher Dive?, 2020) Firstly, I obtained images of both the Kingfisher and Gannet skeletons from Skull Site, a website with several high-quality images and their measurements. I then imported these images into a 3d computer-aided design (CAD) software called FreeCAD, then traced them from the top of the eye socket to the tip of the beak and revolved around the y-axis to create a conical shape that matched the Kingfisher and Gannet skulls. (Appendix A) I replicated this process with seven control models after existing nose cones: SpaceX's Starship and Falcon 9, NASA's Space Launch System (SLS) rocket, and Space Shuttle, and three basic nose cones: Ogive, Parabolic, and Conical. All of the models were then imported into SimScale, a cloud-based engineering simulation platform trusted by professionals worldwide, offering validated CFD used by leading automotive, aerospace, and manufacturing companies and researchers. Firstly, to test for drag, I applied an incompressible simulation, which can help predict aerodynamic forces and airflow patterns, assuming constant air density, to assess the design for stability and drag reduction. After running the simulation, SimScale outputs the Force in Newtons (N) that opposes the rocket (drag). With this, the drag coefficient was calculated using the formula $C_d = \frac{f_d}{\frac{1}{2}\rho v^2 A}$, where C_d is the coefficient of drag, f_d is the force of drag in N, ρ and is the air pressure set at 1.1196 kg/m³, v is velocity in m/s, and A is area in m². The models were tested at 3 different velocities, 500 m/s, 1000 m/s, and 2000 m/s, for 3 different trials to ensure the accuracy of the results.

Furthermore, turbulence was found from the Turbulent Kinetic Energy (m²/s²) using the same incompressible simulation, and no calculation was needed. Digital aerodynamic simulation is a proven method by the researchers at Tsinghua University in their aforementioned study *Design and Analysis of Biomimetic Nose Cone for Morphing of Aerospace Vehicle*, where they tested their nose cone rocket design with a digital simulation. Their study demonstrated that physics-based simulations effectively model airflow, pressure distribution, and structural integrity under different conditions, providing accurate insights into how a nose cone design performs before physical testing. This research justifies the use of computational simulations to analyze rocket nose cones, as it allows for precise testing of drag reduction and stability in a controlled environment. Other examples of physics simulators being used to test rockets and aerodynamics include NASA's Computational Fluid Dynamics simulations, which have been critical in designing the Space Launch System (SLS), and SpaceX's use of physics-based modeling to optimize the aerodynamic efficiency of their Falcon rockets. (Rogers et al., n.d.) These simulations help engineers refine designs, reduce costs, and improve performance before conducting real-world tests, proving their reliability in aerospace engineering. Moreover, by using a credible and accurate physics simulator the data gathered is valid and accurate and answers the research question by directly finding the values of drag, temperature distribution, and stability to assess if the nose cones would be fit for space travel.

Results

The results of this study provide quantitative data on the aerodynamic performance of biomimetic nose cone designs inspired by Kingfishers and Gannets. By analyzing drag and stability through CFD simulations, this study evaluates how effectively these biologically inspired nose cones mitigate key challenges faced by rockets during flight. The CFD simulations conducted in SimScale provided detailed airflow patterns and pressure distributions. By comparing the biomimetic nose cones with existing rocket nose cone designs, such as those used by SpaceX and NASA, the study determines whether the natural adaptations of Kingfishers and Gannets translate into improved aerodynamic efficiency for space travel. The following section presents the collected data, highlighting key differences between the tested models in drag coefficients and stability metrics.

The drag results reveal how the biomimetic nose cone designs compare to conventional rocket nose cones in terms of aerodynamic efficiency, providing insights into their potential for reducing air resistance and improving fuel efficiency during flight. The lower the coefficient of drag value, the better the nose cone is at reducing drag, and vice versa for a high number.

Conical Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-556	0.836
Trial 2	1000	-2240	0.842
Trial 3	2000	-9022	0.843

Figure 4: Conical Drag Table

The Conical Nose Cone simulation yielded an average drag coefficient of 0.84 across the three tested velocities. Since drag coefficients typically range from 0 to 1, this relatively high value indicates significant air resistance, which can negatively impact fuel efficiency, stability, and heat distribution. The conical shape is less effective at smoothly redirecting airflow, leading to increased drag. This resistance opposes the rocket's motion, requiring greater thrust to maintain speed and potentially causing structural and thermal challenges during flight.

SLS Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-584	0.878
Trial 2	1000	-2356	0.887
Trial 3	2000	-9442	0.886

Figure 5: SLS Drag Table

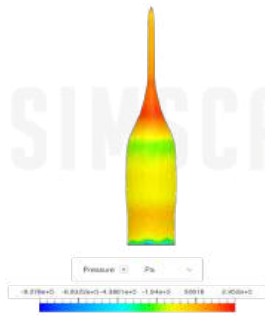


Figure 6: SLS Pressure Image

The Space Launch System rocket had a high drag coefficient value of approximately 0.883 and a large pressure distribution variation as shown in Figure 6. The spike and the body of the nose cone meet, and there is a large buildup of pressure, causing the reddish color, which depicts high pressure. This is caused by the sharp aerospike that displaces the air very well, but it does not have a good transition into the rest of the nose cone and causes a buildup of air particles pushing against the rocket, causing drag.

Parabolic Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-734	1.11
Trial 2	1000	-1806	1.61
Trial 3	2000	-11830	1.113

Figure 7: Parabolic Drag Table



Figure 8: Parabolic Pressure Image

The Parabolic Nose Cone exhibited one of the highest drag coefficients, exceeding the typical range of less than 1, indicating its inefficiency in minimizing drag and displacing air. This is further evident in the pressure distribution map, Figure 7, which shows an unusually high-pressure concentration at the top where air first contacts the rocket. Due to its blunt shape, the parabolic nose cone experiences significant air resistance, leading to excessive drag as the force of the air pushes against the rocket, disrupting smooth airflow and pressure distribution.

Ogive Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-159.79	0.243
Trial 2	1000	-643	0.242
Trial 3	2000	-2594.9	0.243

Figure 9: Ogive Drag Table

Unlike the previous nose cone designs, the Ogive Nose Cone excelled at minimizing drag, achieving an average drag coefficient of 0.243. This result is expected, as the ogive's sharp, streamlined shape allows it to efficiently cut through the air, smoothly redirecting airflow and reducing resistance. By minimizing the force of drag acting on the rocket, the ogive design enhances aerodynamic performance, consequently improving fuel efficiency and stability during flight.

Starship Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-304	0.456
Trial 2	1000	-1201	0.452
Trial 3	2000	-4823	0.453

Figure 10: Starship Drag Table

Space Shuttle Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-317	0.476
Trial 2	1000	-1243	0.468
Trial 3	2000	-5003	0.47

Table 11: Space Shuttle Drag Table

The Starship and Space Shuttle exhibit similar drag coefficient values, ranging from approximately 0.45 to 0.47, corresponding to their comparable aerodynamic designs—short, flat, and rounded. While this shape allows for a smooth airflow transition, it is too blunt to minimize

drag effectively. As a result, the shape generates a significant opposing force against the rocket, reducing overall aerodynamic efficiency. This moderate drag coefficient indicates that while the design balances structural stability and airflow management, it is not optimized for reducing resistance during high-speed travel.

Falcon 9 Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-472	0.709
Trial 2	1000	-1870	0.703
Trial 3	2000	-7552	0.71

Figure 12: Falcon 9 Drag Table

Furthermore, the SpaceX Falcon 9 rocket nose cone has a relatively high drag coefficient, primarily due to its shape. The design features a Starship-like top combined with a more rectangular structure, which, similar to other high-drag nose cones, struggles to efficiently cut through the air. This results in suboptimal pressure distribution and increased aerodynamic resistance. The less streamlined form contributes to greater drag, making it less effective at reducing opposing forces during flight compared to more aerodynamically optimized nose cone designs.

Gannet Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-117	0.266
Trial 2	1000	-674	0.253
Trial 3	2000	2716	0.255

Figure 13: Gannet Drag Table

Alternatively, the biomimetic nose cones performed exceptionally well in the simulations, with the Gannet-inspired design achieving a drag coefficient of just 0.258—the second lowest after the ogive nose cone. This result aligns with the Gannet’s natural ability to dive at high speeds, as its streamlined shape allows it to cut through and displace air particles efficiently. The aerodynamic advantages of this design were reflected in the simulations, demonstrating its potential for reducing drag and improving overall flight performance.

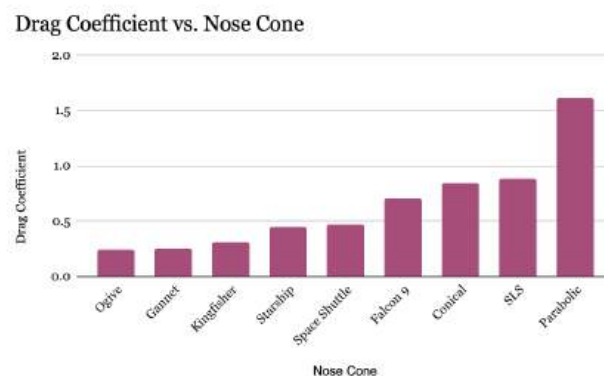
Kingfisher Nose Cone	Speed (m/s)	Total Y Force (N)	Drag Coefficient
Trial 1	500	-204	0.307
Trial 2	1000	-815	0.31
Trial 3	2000	-3287	0.308

Figure 14: Kingfisher Drag Table

The Kingfisher-inspired nose cone also demonstrated strong aerodynamic performance, averaging a drag coefficient of 0.308. This result aligns with the kingfisher's remarkable ability to cut through both air and water with minimal resistance when hunting for prey. Its sharply sloping, conical shape allows air to flow smoothly around it, reducing turbulence and minimizing the opposing force acting on the rocket. This efficiency in air displacement suggests that the kingfisher-based design is well-suited for reducing drag in high-speed aerospace applications.

After simulating each nose cone in SimScale with an incompressible simulation and calculating the drag coefficient, each model can be ranked from best to worst at reducing drag.

Figure 15: Drag Graph



Ultimately, the Ogive nose cone did the best at reducing drag with a coefficient of 0.243. However, the biomimetic models were very close behind, with the Gannet having a coefficient average of only 0.258, which is only a .016 difference. The Kingfisher also had low drag coefficients at an average of .308. The other nose cones start to become less aerodynamic with SpaceX's Starship and NASA's Space Shuttle at 0.45 and 0.47, then .703 for the Falcon 9 and .887 for NASA's SLS, and a drastic 1.16 for the parabolic nose cone. These results are further depicted in their pressure distribution maps in Figure 8; the air has a difficult time being displaced around the nose cone, leading to a greater force acting upon the rocket to create drag. Similarly, the SLS nose cone has a large range in pressure distribution, which can lead to structural stress, cause turbulent airflow, and increase drag, reducing the rocket's efficiency and performance. On the other hand, the biomimetic and Ogive nose cones slice the air with greater precision, distributing the air evenly to reduce drag.

The turbulence results reveal how the biomimetic nose cone designs compare to conventional rocket nose cones in terms of aerodynamic stability, providing insights into their potential for minimizing turbulent airflow and reducing aerodynamic drag. Lower turbulent kinetic energy (TKE) values indicate smoother airflow with less turbulence, contributing to improved stability and efficiency. In contrast, higher values suggest increased turbulent fluctuations, which may lead to greater drag and potential structural stress.

Conical Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1,224
Trial 2	1000	4560.44
Trial 3	2000	1730.7

Figure 16: Conical TKE Table

With its simplistic design, the Conical Nose Cone successfully reduces turbulence, with a TKE value of only 4560.44 m²/s² at 1000 m/s. This is because the shape of the nosecone is flat and minimalist, allowing the air to flow across it uniformly, which means minimal disruptions in its stability. However, at higher speeds, such as at 2000 m/s, the conical nose cone does not perform well, dropping to more of an average value for TKE at 1730 m²/s².

Parabolic Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1238.75
Trial 2	1000	4549.04
Trial 3	2000	16317

Figure 17: Parabolic TKE Table

Unlike its ability to reduce drag, the Parabolic Nose Cone excels at reducing turbulence. With its blunter, rounder shape, the air can move around the nosecone cleanly, not disrupting its stability, although it compensates for its low TKE with higher drag values.

Ogive Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1409.7
Trial 2	1000	4940.14
Trial 3	2000	18618

Figure 18: Ogive TKE Table

The Ogive Nose Cone is highly effective at reducing drag due to its smooth, streamlined shape, but performs only moderately in turbulence control, with a TKE value of 4940.14 at 1000 m/s. This is because, while its gradual curvature maintains smooth airflow, it still generates

turbulence as flow separates toward the rear. Its design strikes a balance between drag reduction and aerodynamic stability, making it ideal for high-speed applications where minimizing drag is the priority.

SLS Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1618.6
Trial 2	1000	6090.87
Trial 3	2000	20229.8

Figure 19: SLS TKE Table

Furthermore, the SLS nose cone is ineffective at reducing turbulence, with a high TKE value of 6090.87 at 1000 m/s. Its design likely causes early flow separation and increased turbulent wake, leading to greater instability. While it may prioritize other aerodynamic factors, its high turbulence makes it less ideal for applications requiring smooth, stable airflow at high speeds.

Starship Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1581.7
Trial 2	1000	5653.2
Trial 3	2000	20729.6

Figure 20: Starship TKE Table

Similar to its performance in drag reduction, the Starship nose cone is ineffective at minimizing turbulence, with a TKE value of 5653.2 at 1000 m/s. Its design likely causes significant flow separation and turbulent wake, making it less optimal for reducing aerodynamic instability at high speeds.

Space Shuttle Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m ² /s ²)
Trial 1	500	1818.79
Trial 2	1000	6559.96
Trial 3	2000	24204.3

Figure 21: Space Shuttle TKE Table

The Space Shuttle nose cone performs the worst in turbulence reduction, with a TKE value of 6559.96 at 1000 m/s. Its design likely causes early flow separation and intense turbulent wake, leading to higher aerodynamic instability compared to other nose cone shapes.

Falcon 9 Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m^2/s^2)
Trial 1	500	1736.01
Trial 2	1000	6636.26
Trial 3	2000	21412.9

Figure 22: Falcon 9 TKE Table

Similarly, the Falcon 9 nose cone is ineffective at reducing turbulence, with a high TKE value of 6636.26 at 1000 m/s. Its design likely leads to significant flow separation and turbulent wake, resulting in greater aerodynamic instability at high speeds.

Gannet Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m^2/s^2)
Trial 1	500	1259.79
Trial 2	1000	4590.33
Trial 3	2000	171176.8

Figure 23: Gannet TKE Table

On the other hand, the Gannet nose cone excels at reducing turbulence, with a relatively low TKE value of 4590.33 m^2/s^2 at 1000 m/s. This efficient design helps maintain a smoother airflow, reducing turbulence and promoting greater aerodynamic stability compared to other nose cone shapes. The Gannet nose cone ranks 4th best overall at both 500 and 1000 m/s and 3rd best at 2000 m/s, demonstrating its consistent performance across different speeds. Its shape likely strikes an optimal balance between drag reduction and turbulence control, making it highly effective in high-speed applications.

Kingfisher Nose Cone	Speed (m/s)	Turbulent Kinetic Energy (m^2/s^2)
Trial 1	500	1192.25
Trial 2	1000	4258.15
Trial 3	2000	15523.8

Figure 24: Kingfisher TKE Table

Finally, the Kingfisher nose cone was the most effective out of all the other nose cones at all three speeds. This is because its unique design, inspired by the streamlined shape of the bird, allows for minimal flow separation and smooth airflow, which greatly reduces both drag and turbulence. The Kingfisher's sharp, tapered nose and smooth, gradual curve promote optimal pressure distribution and minimize the turbulent wake, ensuring superior aerodynamic stability. As a result, it consistently performs better across various speeds, offering the best balance between drag reduction and turbulence control, making it the most efficient choice for high-speed flight.

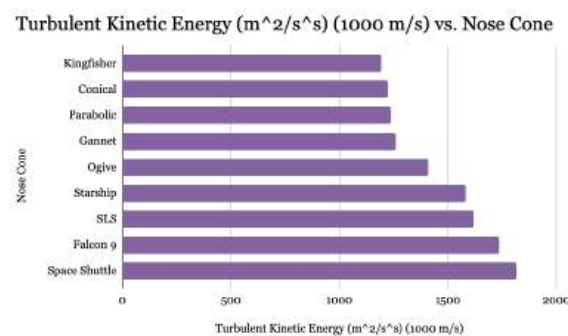


Figure 25: TKE Graph

The results of the turbulence kinetic energy analysis demonstrate the significant impact of nose cone shape on aerodynamic stability. The Kingfisher-inspired nose cone consistently outperformed all other designs, achieving the lowest TKE values across all tested speeds. This suggests that its streamlined shape minimizes flow separation and turbulent wake, leading to greater stability during flight. The Gannet-inspired nose cone also performed well, ranking among the top designs for turbulence reduction while maintaining low drag coefficients. These findings highlight the importance of biomimetic design in optimizing aerodynamic performance.

In contrast, conventional nose cones such as those used on the Space Shuttle, Falcon 9, and SLS rockets exhibited high turbulence values, indicating greater aerodynamic instability. Despite its effectiveness in reducing drag, the ogive nose cone displayed moderate turbulence levels, suggesting a tradeoff between drag minimization and flow stability. The findings of this study are supported by a similar investigation, Repeated Evolution of Drag Reduction at the Air-Water Interface in Diving Kingfishers, in which researchers used computational fluid dynamics to analyze the hydrodynamic performance of different kingfisher species. Their results aligned closely with those of this study, reporting “that [the Pied kingfisher] experienced a lower drag force of 2.27 N compared to [Laughing kookaburra] at 6.86 N, despite similar drag coefficients” of around .23. " (Crandell, Howe, & Falkingham, 2019) Additionally, a study analyzing nose cone aerodynamics, researchers found that a tangent-ogive nose cone had a drag

coefficient of approximately 0.26 at Mach 1.1 and a 10° angle of attack, concluding that “tangent ogive shapes produce relatively low drag in supersonic regimes” (Nakka, n.d.). This closely aligns with the drag coefficient of 0.24 calculated in my own study, supporting the validity and accuracy of my results. Overall, the results suggest that biomimetic designs inspired by kingfishers and gannets effectively reduce both drag and aerodynamic instability, which is crucial for improving rocket efficiency, stability, and overall effectiveness.

Implications

The results of this study have significant implications for further research and effects on real-world applications. The research findings showed the effectiveness of the biomimetic nosecones using Kingfishers and Gannets. They are shown to be especially effective at reducing drag and maintaining stability. Further research could create more extensive tests on the effectiveness of the Gannet and Kingfisher nose cones. This could be done by creating physical experiments, such as wind tunnel testing, that would further validate the data collected from the simulation and expose the designs to more accurate and intensive trials.

Furthermore, the research could also test the nosecone's ability to withstand high-temperature buildup that would affect the nosecone during reentry. This would provide a more comprehensive analysis that includes another problem that rocket nosecones face during flight, and that the Gannet and Kingfisher designs must be able to withstand. Alternatively, this research highlights the ability of nature to inspire engineering ideas, such as the aerodynamic nosecones from diving birds, to create unique and effective solutions. This can help engineers look to the natural world for inspiration and use those ideas not just on aerospace but in any kind of engineering problem.

Limitations

Despite the success of the results for the Gannet and Kingfisher nosecones in reducing drag and turbulence, there are limitations. Firstly, the data that was collected is from a physics simulation. Although it is credible software that is used by engineers and businesses from around the world, it is still a hypothesis of what will happen if these nosecones are ever used. This means that for the data that was collected, another study must continue the research and apply the findings to a real-world test, such as a wind tunnel. Additionally, this study only looked at the variables of drag and stability. Although they are some of the most important, further research would need to be done on other variables in flight such as shock wave formation or temperature distribution which could severely impact the success of the nose cone during flight and therefore show other reasons as to why or why not the nose cone designs would be effective. Regardless of these limitations, the data collected in the study is still valid and applicable to the implications.

Conclusion

Moreover, this study highlights the potential of biomimetic nose cone designs to enhance aerodynamic efficiency by reducing drag and turbulence. Inspired by the streamlined shapes of

Kingfishers and Gannets, these designs demonstrated strong performance in computational simulations, closely rivaling or even surpassing some conventional nose cone shapes. The findings suggest that incorporating biological principles into aerospace engineering can lead to more effective solutions for improving rocket stability and fuel efficiency.

Beyond the immediate aerodynamic benefits, this research underscores the broader value of nature-inspired innovation in engineering. Engineers can develop optimized designs that align with real-world performance needs by studying how organisms have evolved to navigate air and water efficiently. Future work should focus on experimental validation and expanded testing to explore additional variables such as thermal resistance and structural integrity. Ultimately, this study contributes to the growing field of biomimicry by addressing the gap in understanding how nature-inspired designs can enhance rocket performance, specifically by demonstrating that a nose cone modeled after kingfishers and gannets can reduce drag and instability, demonstrating that natural adaptations can drive technological advancements in space exploration and beyond.

Works Cited

- Alcedo atthis (Common Kingfisher) – skullsite. (n.d.). Skullsite.
<https://skullsite.com/skullpage/alcedo-atthis-common-kingfisher/>
- Ayre, M. (n.d.). Biomimetics Applied to Space Exploration.
https://www.esa.int/gsp/ACT/doc/BIO/ACT-RPR-BIO-2004-Biomimicry_Space.pdf
- Basics of Rocket Aerodynamics Optimization. (n.d.).
<https://www.apogeerockets.com/education/downloads/Newsletter629.pdf>
- Build Your Own Wind Tunnel. (n.d.). Wwww.grc.nasa.gov.
<https://www.grc.nasa.gov/www/k-12/WindTunnel/build.html>
- Choi, J., Hwang, J., Jeong, Y., Park, J. M., Lee, K. H., & Hong, J. W. (2015). Biomimetics: forecasting the future of science, engineering, and medicine. *International Journal of Nanomedicine*, 10, 5701. <https://doi.org/10.2147/ijn.s83642>
- Crandell, K. E., Howe, R. O., & Falkingham, P. L. (2019). Repeated evolution of drag reduction at the air–water interface in diving kingfishers. *Journal of the Royal Society Interface*, 16(154), 20190125. <https://doi.org/10.1098/rsif.2019.0125>
- Engineering, I. (n.d.). How kingfisher inspired bullet trains? Interesting Engineering.
<https://interestingengineering.com/videos/ep-1-how-kingfisher-inspired-bullet-trains>
- Hall, N. (2023, November 20). Rocket Aerodynamics | Glenn Research Center | NASA. Glenn Research Center | NASA.
<https://www1.grc.nasa.gov/beginners-guide-to-aeronautics/rocket-aerodynamics/>
- HAN, J., HUI, Z., TIAN, F., & CHEN, G. (2020). Review on bio-inspired flight systems and bionic aerodynamics. *Chinese Journal of Aeronautics*, 34(7).
<https://doi.org/10.1016/j.cja.2020.03.036>
- Kingfisher Diving - How Fast Does A Kingfisher Dive? (2020, July 22). Bird Baron.
<https://www.birdbaron.com/kingfisher-diving/>
- Rogers, S., Dalle, D., & Chan, W. (n.d.). CFD Simulations of the Space Launch System Ascent Aerodynamics and Booster Separation. Retrieved March 3, 2025, from
<https://ntrs.nasa.gov/api/citations/20190001990/downloads/20190001990.pdf>
- Sia, J. S. (2021, June 24). Rocket Physics, the Hard Way: Re-entry and Hypersonic Flight. The Mars Society of Canada.
<https://www.marssociety.ca/2021/06/24/rocket-physics-the-hard-way-re-entry-and-hypersonic-flight/>
- Space, A., & Sesi Kottapalli. (2017). Boeing Smart Rotor Full-scale Wind Tunnel Test Data Report Nasa/Tm-2016216048. Createspace Independent Publishing Platform.
- Ten Bio-inspired Aerospace Innovations: Based on Biomimicry of Insects, Birds and Animals – EDI Weekly: Engineered Design Insider. (n.d.). Wwww.ediweekly.com.
<https://www.ediweekly.com/ten-bio-inspired-aerospace-innovations-based-on-biomimicry-of-insects-birds-and-animals/>
- Wind Over Its Wing: NASA's X-66 Model Tests Airflow - NASA. (2025, February 5). NASA.
<https://www.nasa.gov/centers-and-facilities/armstrong/x-66-model-tests-airflow/>

Zhao, J., Yan, S., Deng, L., Huang, H., & Liu, Y. (2017). Design and analysis of biomimetic nose cones for morphing of aerospace vehicles. *Journal of Bionic Engineering*, 14(2), 317–326. [https://doi.org/10.1016/s1672-6529\(16\)60400-6](https://doi.org/10.1016/s1672-6529(16)60400-6)
(2025). Nakka-Rocketry.net.
https://www.nakka-rocketry.net/RD_nosecone.html?utm_source=chatgpt.com

“His Family Tree Recedes”: A Personal Narrative Analysis Discussing Dementia Caregiving Themes, Strategies, and AI Implementation By Noel Yi

Abstract

As the population of older adults increases, there are not enough dementia caregivers in the healthcare workplace, resulting in an increased pressure on close family members to provide care themselves. Today, dementia can significantly affect family relationships, often leading to their deterioration as the illness impairs an individual’s cognitive processes. This makes communication, perception, and general connection harder to navigate.

This paper addresses conditions and stressors that push family members into the obligation of being primary caregivers for people with dementia (PWD). By drawing on personal familial experiences and scholarly literature, I will present strategies to mitigate the impact dementia has on family caregivers and their relationships with relatives coping with dementia. The paper also presents how these interpersonal strategies will offer a more robust understanding of dementia treatment, shape policies, and cut loose caregiver isolation and stigma via public education at schools and community centres.

I conducted interviews across four generations of my family to get a comprehensive view of their diverse caregiving experiences and perspectives. In coding keywords, the most relevant themes for family caregivers were filial piety/seniority, caregiver burnout, relational conflict, affection, coping, and lack of support. I apply data collection and content analysis of real-world online forums designed for dementia caregivers.

There is a rising possibility in the potential of AI in home intervention and in healthcare institutions. A change in policy, in conjunction with dementia-specific education programs are necessary to promote effective communication between relatives and make knowledge accessible for caregivers.

Introduction

Many of the encounters involving my paternal grandpa with Alzheimer’s disease and his caregivers were negative. They were often a recounting of how my grandpa berated another caregiver until they were forced out the door. Whenever I heard that we had a new hire, I knew it was only a matter of time until they also reached their breaking point and the scale tipped out of balance, patience dissipating into pure frustration. Our nannies and caregivers were practically sent out the moment they stepped through the door and before they could say “Hi.”

The combination of senior age and dementia does not fare well, and my grandpa was no exception. As my grandpa became increasingly fragile, my family decided against sending him to a memory care facility or senior home because the quality of care was too unreliable. They grimaced at relatives’ anecdotes and general media that exposed cases of mistreatment and neglect, leaving my paternal uncle and grandma to take on the brunt of caretaking.

Still, my uncle was the primary caregiver of grandpa’s dementia, as my grandma was vulnerable to health concerns. My uncle began working increasingly from home, leaving his

office life behind. During an interview, he stated, “This would not have been possible if my work was not running a business with my son. If I didn’t have the privilege of flexibility, I certainly would have had an additional layer of financial stress while caregiving...In fact, I wouldn’t have been as available as I was.” His assessment led me to question who might have taken on the caregiving duty if my uncle was not present.

Due to living in close proximity to one another, my uncle became the “supervisor” of my grandpa’s house. Some of his responsibilities included: paying the bills, accompanying doctor’s appointments, buying their medicine. My uncle also felt the strong obligation to demonstrate filial piety. Based in Confucian thought, filial piety is the reverence one has for one’s ancestors and parents (“Filial Piety”). While his sisters and brothers all pitched in to cover costs for my grandpa’s care, this uncle offered his emotional labor, checking in with my grandparents’ morning and night, with only his afternoons left for work endeavours. These examples highlight the burdens of caretaking and emphasize the sacrifices that my uncle had to make.

In my uncle’s case, while he loved my grandpa dearly, the constant resentment grandpa held for him negatively impacted his own mental health. When asked if there were any caregiving strategies he adopted, he said, “There is no strategy, I just grinned and bore it. You can’t get your point across to him in arguments. Grandpa always thinks that he is right and that he hasn’t made a mistake. And he’s my parent, so I just suppress my burnout.” In my uncle’s opinion, this pattern of self-suppression was manifested in his later diagnoses of anxiety and clinical depression. He still takes anti-depressants to this day, despite my grandpa not being with us anymore.

In dementia caregiving, one may allocate all their energy to being a necessary resource for a person with dementia (PWD), only to be repeatedly told that they do not want your help. If my uncle overstayed my grandpa’s welcome, proposing to sleep overnight under their roof to offer 24 hour assistance, grandpa would rush him out like any other hired caregiver. In contrast, my grandma stayed relatively emotionally detached from grandpa’s hurtful words and spontaneous aggression. She did not feel like she had caregiving strategies either, saying, “I can give my best attempt at telling him what to do, but if he doesn’t listen, he just doesn’t. What can I do? So I don’t let it affect me psychologically.”

While a PWD may be the only one displaying weakened physical health, dementia affects the PWD’s entire family, across generations. I have interviewed five relatives from four generations of my paternal family to gain a comprehensive look at my grandpa’s four and a half years of dementia and to juxtapose all their narratives to see what it reveals about what should be done to improve caregiver relations with PWD and other family.

My interviews made it clear that both my grandma and uncle felt like they did not have strategies to alleviate the burden of dementia caregiving. There is an insufficient amount of targeted support for these primary caregivers, and it can leave them feeling isolated or questioning whether they are providing the right kind of care. As the global population inevitably ages, dementia is becoming one of the leading causes of death in the world. There must be more attention to improving the state of the relationships between care receiver and caregiver.

Although there are organizations like Alzheimer's Disease International and Alzheimer's Association that educate people about dementia, there is the need for more awareness and actions in regards to correcting the stigma of dementia, a stigma that can be seen in modern media, which too often unfairly and superficially portrays dementia patients and their stories.

In this paper, I will address the conditions and stressors that make it necessary for family members to become the primary caregivers for PWD. Drawing on personal familial experiences and scholarly literature, this paper presents strategies to mitigate the impact dementia has on family caregivers and PWD relationships. It also suggests how these interpersonal familial strategies might offer a more robust understanding of dementia treatment, shape policy, and education.

Care, Complexity, and Context

As the average human lifespan increases, dementia becomes an increasingly relevant topic. The most recent World Alzheimer Report 2024 from Alzheimer's Disease International reports that "every three seconds, someone, somewhere, develops dementia" (Alzheimer's Disease International et al. 8). According to the World Health Organization, this number will reach 78 million in 2030 and 139 million in 2050, meaning it nearly doubles every 20 years ("Dementia Statistics"). Since the disease will only become more widespread, current anthropological and medical research communicates urgency in implementing caregiver strategies.

Socioeconomic status (SES) plays its role in the way an individual's dementia can impose a burden on caregivers through a heavy economic toll. An estimated USD 2 trillion will be allocated globally to dementia by 2030 (Wang et al.). Rehabilitation efforts, medication, and home care services are expensive. Senior homes with specialized long-term memory care units are often out of reach for most families. Additionally, education level and poor environmental conditions are risk factors for dementia. Several studies have reported that low SES is associated with many adverse health outcomes (Wang et al.). Despite the relative lack of SES-related case statistics, this is a well-established connection in current dementia literature. Some reviews and meta-analyses have discovered that each additional year of education for an individual is associated with a 7 percent lower risk of dementia (Xu et al.). This correlates with the idea Huang et al. discovered that more mental exercise and experience in complex occupations were also associated with a lower risk of dementia.

Treating my family's circumstance with my grandfather progressing dementia as a case study, I can affirm that we have been privileged to have family available to take care of our grandfather. For other households, it may be necessary to quit work in order to properly allot time to caring for their affected relatives. This can leave a deficit in the regular income the whole family makes, thus placing them in a worsened economic state, which may then lead to lack of therapy and medical care. However, I also wish to emphasize that while money will always matter when it comes to external help services, with dementia, there is no cure yet. As my grandmother has noted, "dementia does not care how much money you have. All cases end the

same because there is no treatment.” In Dalian, China, which is where my family is from, my uncle mentioned that there are senior living homes, but institutionalized care there does not target the dementia itself; rather, they focus on solving dementia symptoms. Dementia is clearly relevant and caregivers urgently want reliable therapy for this illness.

A needs assessment survey of youth caregivers gathered the popular sentiment that, “doctor’s don’t know anything about dementia” (Nichols et al. 26). The assessment reported a consensus from participants about the lack of support from healthcare professionals. This is likely because youth caregivers don’t feel any comfort from accompanying their PWD to hospitals—there is nothing they can offer that will make dementia an easier psychological and physical battle for their loved ones.

More than 90% of respondents and caregivers said that they would be motivated to get a diagnosis if there was a disease-altering treatment (Alzheimer’s Disease International et al. 9). With a correct diagnosis, people with dementia and their families can open the door to support groups, information, and feel part of a community, enabling them to receive better care in a timely fashion. Groups can also facilitate access to new life-prolonging treatments and educate people about the ways a diagnosed individual can live in their own home for longer periods of time.

This research project is a way for me to understand and relay the diverse experiences of dementia. As Alzheimer’s Disease International writes, “the voices of people with dementia will break down barriers—and this can enhance self-esteem, quality of life, promote social inclusion, and improve overall wellbeing” (Alzheimer’s Disease International et al. 12). My family’s experience is an opportunity for me to do my part and enter the discussion about dementia awareness. As anthropologist Janelle S. Taylor says, the stigma of living with dementia is often undeserved and exaggerated unfairly in the media, in seemingly “gothic and zombie stories” (Taylor 324). Taylor’s experience of caregiving for her mother with dementia provides a valuable correction to media portrayals that frame dementia as simplistic, or that play into one cookie cutter “horror” dementia experience. She writes, “What they find much harder to hear, I think, is that I am not a victim, and being around my mother is not a nightmare or a horror. She is not “dead,” she is not “gone,” and she is not just a “body”” (Taylor 324). Taylor offers another perspective of caretaking, showing how experiences of dementia are far more individualized than standardized. She reminds us that some caregivers don’t feel like a victim, and don’t want to be pitied. Taylor critiques the notion that PWD are just bodies, and that those around caregivers always ask, “well, does she recognize you?” She identifies the societal tendency to connect personhood to an ability to communicate one’s own thoughts and introduces a discussion relating to the ethics of such a link. Although dementia can rupture relationships, some still cherish the non-verbal communication in interactions with PWD.

Taylor also recognizes that her caregiving experience was relatively fortunate compared to what others might go through, which also helped me reflect on the balance between having a relative who is unpredictable while also loving them and wishing them better. Dementia is a sensitive topic, and it needs to be handled with particular attention to the PWD’s needs, habits,

and traits. Disregarding that fact may lead to societal isolation from caregivers. There is a negative stereotype society places on dementia, but it is more complex than what one might think. I suggest readers consider this diversity of experience.

Dementia Caregiving Behind Closed Doors

In many households, dementia care may look like small practices the caregiver gradually employs as they spend more time with the relative with dementia. This might be learning to go ahead with mistaken stories or facts the PWD brings up in conversation, agreeing with what they want to avoid verbal, or physical conflict. This may also mean applying “life story work” by sitting down and reminiscing about old experiences to share a moment with the PWD and induce cognitive activity, which can stimulate cognitive processes—delaying cognitive decline. The family might also seek out support groups from dementia organizations. However, it was observable that there is a lack of support systems for youth caregivers (18 years and under) in particular. Perhaps this is because of liability reasons, but youth caregivers should have external help available to them, if not more than other age groups because of the sensitivity of brain growth and mental illness stemming from caretaking like anxiety and depression.

If the family spreads across multiple generations under the same roof, sometimes the members will take turns supervising or specific members will bear the brunt of caretaking responsibilities. This can result in family alienation, where specific family members are depended on while others withdraw, leading to social isolation and resentment. Sometimes, this occurs because the person with dementia does not accept anybody else as someone he knows and recognizes—making those isolated few the most well-trusted for the patient. Other cases of isolation may occur if other family members choose to monetarily imburse the caregiver if they don’t have enough time or are unmotivated. In a study aiming to explore and develop the process of self-empowerment for primary caregivers responsible for caring for elderly dementia, an interviewee mentioned, “Every time I asked him [my younger brother] to take care of mother over the holidays, he would take a long time to give me an answer. He might think that he doesn’t need to take this responsibility because he has always paid me” (Che et al. 215). Neglecting a primary caretaker’s needs can cause family alienation, even if the neglect is accompanied by money.

Extensive security measures to protect relatives with dementia are also used in households, as well as senior homes: “We usually just close our front door without locking. Now, we must lock the door with keys for mixed locks” (Che et al. 215). At a senior home in Vancouver, visitors and seniors have to buzz in via call with the front office. I didn’t notice much about it at the time, but then I recalled that one day my grandfather was at home in our complex and he wandered out the door, took the escalator down, and strolled around in the courtyard, completely lost. Until we heard him arguing with a staff downstairs, we were petrified that he was no longer in the home. That’s when my uncle began spending most of his hours of the day with him. Locks and buzz-in doors can be a necessary precaution taken to protect people with dementia. Another senior home I visited had double sliding doors that closed for a significant

time before allowing you to go through—possibly delaying a confused person from just exiting in the middle of the night and potentially ending up in danger.

Some people may not send their relatives to memory care homes because amenities are expensive and the quality of life is unassured. For my family, we didn't trust the institutionalized care system in China enough to send my grandpa to a memory care home. We had heard from friends who knew that they would bind relatives with dementia to beds or insert IVs to calm them down. We preferred to have our grandpa stay in the relative safety of our home. In the end, caretaking for a PWD encompasses all of these components, twenty-four hours a day, seven days a week. It can feel like an overwhelming sense of duty and responsibility, especially with a lack of strategies to cope or help mitigate the rapid loss of a stronger, earlier, relational bond.

Through the Lens of a Granddaughter

As a child, dementia seemed to me a natural part of the aging process. But I heard of friends' grandparents being much more active and engaged in their lives, and I wondered what life might look like for them. While the research in this paper is informed by scholarly literature, everyone's experience with dementia and Alzheimer's is different. The dementia narrative can be hard to quantify, even with statistics of its relevance and percentiles of community confidence that have been reported by organizations. My research offers suggestions to provide some solace or guidance, but its basis lies in my personal narrative and personal connection to the illness. In that way, the lessons I've learned may not be best extrapolated for every other dementia case out there, especially given that my grandpa and his story lived in Dalian, China, so the resources and experience may be regional and country-oriented. The details of dementia have always been a mystery to me, so this research also serves as an extension of my curiosity and genuine interest into reflecting on this period of my life.

Seeing my grandfather progress through 5 years of dementia, from 2019-2024, Alzheimer's is like seeing someone's memory slowly grow dappled. The illness brings a targeted cruelty, it seems like a way nature attempts to extend but also slow down the grief of a loved one passing away. Feeling that anguish every passing day can tend to chip away at a family's morale, just like my uncle experienced. It is debatable whether dementia softens the grief or deepens it, as sometimes, watching over a dementia patient is like being present while their spirit keeps diluting into disorientation. It seemed to me that cognitive regression was no different than a pen running out of ink, or a soap bar becoming more amorphous with every wash.

My grandfather's memory reflected that of a light switch being turned on and off repeatedly; he could remember his three sons and two daughters one day, and then the next day, he couldn't.

He could never quite remember me; perhaps it was because I was among the latest to arrive in the picture of his life. When I returned to Dalian, China every summer, it saddened me to know I was a stranger when I rang the doorbell and waved my hand at him. And a stranger when I sat across from him on the living room sofa. Everytime, my dad would hustle me over and tell me to introduce myself once again, trying to jog his memory.

“Hey grandpa, I’m Noel, I am your youngest son’s daughter” became the quintessential phrase to start off any conversation.

Chinese New Year: “Happy New Year grandpa! I hope you have been resting well and I wish you a healthy body and mind this year.”

Moon Festival: “It’s Noel, grandpa. I hope you feel well-fed with the festival dinner your sons are taking you out to. Have fun.”

After each moment of silence, a grunt and blank eye contact made us feel so distant, even though we were technically so close.

As I grew busier in the United States and Canada, I found myself not able to return to China as frequently. The cozy atmosphere the Dalian household once had was snapped by a tenuous thread, and now it was mentally draining for everyone. It was someplace where I felt immersed in my greater family and my connection to my Chinese culture as well. Western influences have taken a large part of my identity as I’ve grown and spent my time here, so I appreciated having the simple pleasures of folding dumpling sheets over filling with my aunt and playing around with the extra mahjong tiles in their wooden box as my dad and his brothers played and laughed. This home was important to me—and my grandfather’s Alzheimer’s was proof of how one person’s valued presence can induce an entire family network to topple.

I interviewed four generations of my family about my grandpa’s disease progression: my grandma, my uncle, my cousin, and my nephew. They mentioned how one day in 2019 grandpa just changed: he forgot his son’s names, he would cough up phlegm onto the floor, and he would be angrier than normal when someone said he made a mistake.

In 2023, he got radically worse. My grandpa always had a rather short temper, but he would never spew too many insults, as he was the type to keep polished and make sure to give a polite impression to guests and people he didn’t know. When he had Alzheimer’s, everyone would seem like a stranger to him except for my grandma and my uncle. He’d wake up and find a bunch of strangers cooking breakfast in his own kitchen and watching his own tv.

You’d think that, given his personality, he would keep quiet about complaints. And he did, until 2023. Although Frontotemporal Dementia (FTD) is typically more famous for affecting the prefrontal cortex, a region of emotional regulation in the brain, Alzheimer’s can do so as well. His own customs changed, and while I can’t fully speak on his behalf, I believe he would’ve been embarrassed by the way he treated others in his later years. He wasn’t the courteous self he prided himself on. Consequently, my grandpa was made severely prone to irritability and anger. My cousin mentioned how lucky we were that grandpa didn’t make deliberately dangerous actions or decisions. He won’t touch fire or go grab a knife without knowing, and he wouldn’t put hazardous items in his mouth as food. This instilled comfort is us knowing that he still had some instincts in regards to safety.

Admittedly, when I heard this was happening as a kid who didn’t understand dementia, I used to feel ashamed that he would do that and afraid of my grandpa’s emotional regulation – or rather lack of. I wasn’t able to comprehend because I wasn’t given the time and education to –

and this area of support should receive attention, especially in intergenerational households who have dementia patients.

Thematic Analysis

After considering all three of my interviews (grandma, uncle, cousin), six sub-themes have emerged from the transcripts: caretaker burnout, relational conflict, coping strategy, filial piety, lack of support, and affection. I coded out the frequency of the reference of these themes in each separate interview by using keywords and Google Colab. ‘Caretaker Burnout’: depressed, exhausted, tired, burnout; ‘Relational Conflict’: argue, yell, hit, anger, conflict; ‘Coping Strategy’: cope, strategy, manage, align; ‘Filial Piety/Seniority’: duty, obligation, responsibility, owe, eldest; ‘Lack of Support’: no help, lack, unsupportive, alone; ‘Affection’: love, affection. Here were the theme frequencies in interview transcripts of my grandma, cousin, and uncle, respectively. These figures display each theme’s relevance in each interview. I have chosen two overarching themes that touch on all of these: PWD Carrying Habits and Traits from Youth into Senior Life and Inside and Outside Pools of Dementia Caregiving Within Families.

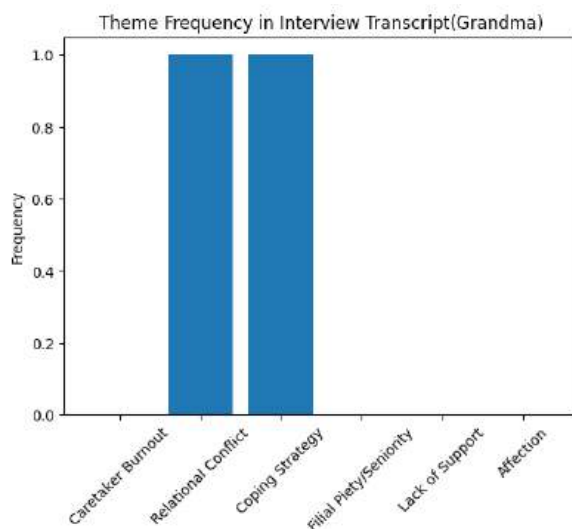


Fig. 1. Frequency of caregiving themes in Grandma’s interview transcript, mentions of relational conflict/coping strategy.

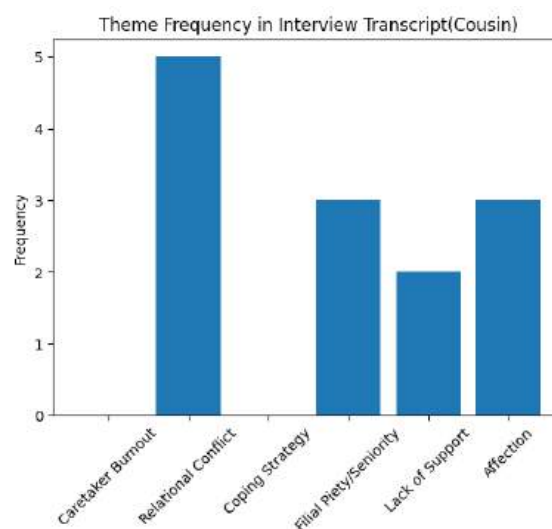


Fig. 2. Frequency of caregiving themes in Cousin’s interview transcript, mentions of relational conflict/filial piety/seniority/lack of support/affection.

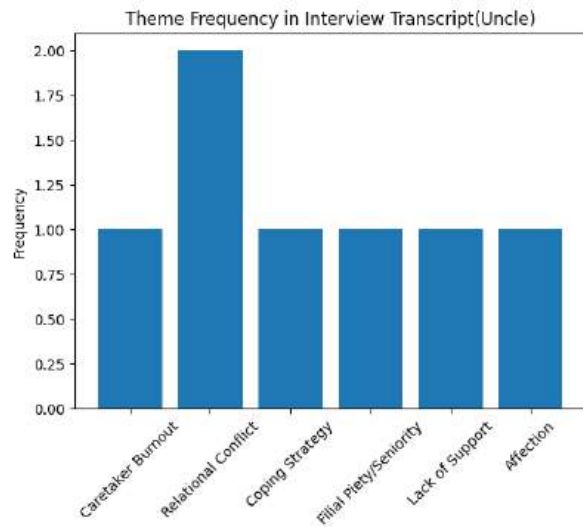


Fig. 3. Frequency of caregiving themes in Uncle's interview transcript, mentions of all themes considered.

As the eldest, my grandpa expected respect. My uncle said that, "Grandpa did not feel empathy when he made a place unclean or hurt someone because he was the eldest." Respect for seniority is a primary part of Chinese customs, and thus is an important contextual point for this paper. Dementia did play a major role in ruining my grandpa's etiquette with those younger than him, but his elevated position as an elder in Chinese tradition also contributed to this personality change.

I. PWD Carrying Habits and Traits from Youth into Senior Life

We often know the incapacity of recalling personally relevant events from the past as a quintessential early symptom of dementia. However, sometimes Autobiographical memory (ABM) can deteriorate differently on a case-by-case basis.

- "In his final years, grandpa didn't recognize his house as his own. When someone mentioned his house, he thought of his childhood house nestled in rural Dalian as opposed to the modern apartment complex he actually lived in. He would have felt more settled in returning to his old neighborhood to see out his last days. He would have liked seeing the same steel pots and ladles he saw as a kid—his mind was stuck in that timeframe. I wish we could've brought him back so he could've finally felt settled, but the friends he refers to have passed, that kitchenware is surely no longer there, and it's just not feasible to pack up to live in drastically worse conditions after we've made our lives here."
- —Cousin, grandpa's grandson (transcribed/translated from interview).

ABM refers to an individual's ability to draw on personal semantics and other episodes to buttress a strong personal identity over time. ABM is anchored in the discovery of the "core recollection network", in and around the brain's hippocampus. Key regions include the

ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex, lateral temporal cortices, lateral parietal cortices, and the hippocampus itself (Figure 1). Damage done to any of the nodes in this core recollection network are said to hurt ABM retrieval (Irish).

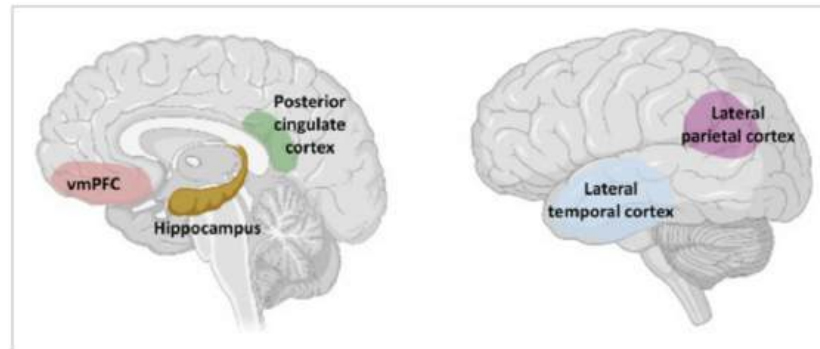


Figure 4. The five key regions of “core recollection network.” Image Source: Irish

According to neuroscientist and neuropsychologist Irish, early studies of ABM in Alzheimer’s Disease (AD) show that individuals’ ability to retrieve recent events was more compromised in comparison to events that are more remote temporally. In other words, episodic memory degraded while personal semantics remained somewhat intact. Consequently, there was a more proficient recall of memories formed in childhood and early adulthood in comparison to memories formed in more recent time periods.

Irish also writes that although earlier studies of ABM favor the standard consolidation model of memory, when memories are said to gradually grow fully independent of the hippocampus—where episodic events are typically stored in short-term memory—and move to neocortical sites—which house long-term memory and thus mark memory consolidation. However, more recent studies have turned to support the Multiple Trace Theory (MTT) instead. MTT argues that the hippocampus is always involved in both storage and recalling long-term events, it’s just that if semantic information is given the time to establish itself into the neocortex sufficiently, that long-term information will survive damage to the hippocampal system (Nadel et al.). This does not mean that remote memories are easier to retrieve, rather, they require as much activation effort as retrieval of recent memories in the hippocampus. In the MTT, the relative intactness of early episodic memory is related to the amount of ‘traces’ laid down prior. With early memories having more reactivations, this means that there are more traces and they are relatively more intact in comparison to episodes that were just encountered or experienced (“Multiple Trace”). As my grandpa entered the final stages of dementia progression, my grandpa’s profile fit anterograde amnesia, as he could not make any new memories (“Anterograde Amnesia”).

My grandpa’s AD experience relates to today’s neuroscience discussion by aligning more with the Multiple Trace Theory, less with the standard consolidation model of memory. This was evident in the way he could not retrieve the recent hippocampus-dependent episodic information of his most recent home. He had a bias towards remote autobiographical recall of events when it

came to thinking of his old childhood house and the experiences linked to it “home”. His memory of the kitchenware and his old friends represented vivid details, offering promise in classifying these memories as being strengthened by multiple traces. In addition, what interested me the most when conducting my interviews was how my grandpa reverted to his old habits. Outside of my current research, I hadn’t heard about this being a possible outcome of AD. Living in old rural DaLian, my grandpa had less food security as a child, because he fell under poorer circumstances. As my cousin said, “One of his habits in childhood was hiding and taking money and food, because he was scared of other people taking. He was taught that resources were never sufficient for everyone.”

My grandpa’s tendency to hide food and become irritated when other people touched or moved it cannot be justified solely by muscle memory, as he stopped this habit before his years of dementia. There was no reason for the reinvokement of this activity as his house had sufficient amounts of food now, but my grandfather’s memory followed the classic temporal gradient of retrograde amnesia. Furthermore, given that a child’s brain is especially malleable up until 25, the need to secure food became a stronger-rooted memory in his brain. It was likely that his associated emotions influenced the strong establishment of the memory by way of emotional enhancement of memory (EEM). This is when the amygdala is especially involved when responding to emotional stimuli of a moment and it results in more attention to this stimuli and thus improves consolidation (Chainay et al.). Connecting this back to the MTT, better encoding can enhance its ‘trace’ and mean better subsequent recall of the event.

Thus while my grandpa’s AD caused him to become more irritated in his entirety, my grandpa’s AD also enhanced personality traits that were already within him as the illness progressed. My grandpa was an orphan who didn’t always feel safe, so skepticism and caution were exercised as his early personality patterns. In early-stage dementia, while he was always skeptical of medical interventions, the two people he trusted at the time, my grandma and my uncle, would end up persuading him. Later, in late-stage dementia, my cousin noted that,

- “At a hospital, he will try for any sort of solution to escape from catheters and IVs. Even though it hurts, he would still go and take it out. He thought that it didn’t feel right that he was attached to the bed. He didn’t trust any procedures and treatment.”

At the same time, one aspect of MTT is that the information in the neocortex is still likely to be affected by harm done to the hippocampus, as these parts of the brain are still used in one pathway of the storage of memory and not completely distinct. This is reflected in my grandpa’s inability to recruit some remote semantic knowledge. Semantic knowledge is a term used to describe all the general concepts and basic information people learn from birth to the present (Kintz and Wright). This includes information like apples being edible, what kitchen tools are used for what purpose, etc. For my grandpa, in his late-stage dementia, my grandma said that,

- He didn't remember me as his wife. He just thought that I was family, a close person, perhaps like a sister. He didn't comprehend restaurant menus and foods. I had to tell him what bok choy was, or some other ingredient that was listed. Similarly, you would see him watching the TV, but he didn't understand what they were referring to or talking about. It was as if he was a kid, learning the basics all over again."

By framing this connection with MTT, I demonstrate that my grandpa's behaviour was not erratic, it has neurological underlying causes. I hope to show work that bridges have illuminated what other family caregivers might see with context and normalizing this experience. Scientific advances can serve as a grounding basis for some families in the same position.

If MTT is right, we shouldn't assume remote memories are safe or spared from degradation, instead we should implement more activities and medical policies that push for utilizing these valuable cues and reinvoked memories, so that more traces can be laid, and cognitive regression can be slowed and mitigated. This also suggests that there may be promise in tuning into environmental life story cues as ways to reconnect relationships between caregivers and PWD. If possible, attaining small memory cues like an object that was prevalent in childhood or early adulthood can make a PWD feel more at ease, reduce distress, and help them connect with their identity.

It is critical that caregivers gain an ample knowledge of a PWD's childhood traits or of the habits the PWD developed earlier in life. Learning about their tendencies, perhaps what they are afraid of internally and have a passion for, can help create a relationship of more trust, bettering the experiences of caregivers and dementia patients. Knowing all this information can help the caregiver have more of an idea of what topics to avoid, what to work on, and what to immerse into conversation about. To attain this information about early experiences, it might be useful to offer the patient an opportunity to write down a few things they can remember about themselves as a kid or young adult. Such a technique can be incorporated into life story work at home. Life story work has always been proven in some studies to be effective, a method to care as a "hands-on" technique for older people. Dementia UK shares that life stories are the "fact file" of a PWD, that includes their interest, backgrounds, values, close connections. It can be in the form of books, collages, video recordings, memory boxes, apps, and personal profiles, shared with other people, including healthcare professionals, family, friends, and carers ("Creating a Life").

One of the problems a PWD may struggle with is telling stories in a sequence that can be understood easily by the listener. Telling a chronological story can be especially difficult given cognitive impairment, and it's a potential explanation for why patients see a receding family tree or shrinking friend community. With that, it becomes difficult to not misattribute names or places, leading to overall stigmatisation and confusion for a listener. That's why Kindell et al. argue that the primary point of life story work is not to focus on the verity of the subject matter, instead it's to unpack and understand why a certain individual chooses to share that story specifically (Kindell et al.).

My analysis shows how habits and traits from youth should be used productively to alleviate caregiver stress or confusion through life story work and framing caregiving in small environmental cues, like bringing up an object that might reminisce about their youth and enhance communication (Subramaniam et al.). It's important to incorporate life story work into care work today, as it can defend against stigmatisation by bringing PWDs an outlet for preserving their identity and also have other people better understand and relate to them. Caregivers should have the access to framing their experiences into a neurological context as well, and equipping them with centralized resources, in addition to the existing forums in use right now can be useful. With more knowledge, caregivers can have more strategies in their toolbox to tackle every-day struggles.

One of the shortcomings of this is that it can only be successful in alleviating stress for caretakers if they have an accessible database to ask questions, where information about all types of dementia are written into lay terms. My relatives had no idea where to access information about dementia and how to care for someone struggling with the disease. By also allowing caregivers to converse amongst themselves on this platform, the online community can demystify each other's uncertainties and limit feelings of isolation, which happens when a caregiver feels like they are figuring everything out alone.

Many support group organizations tend to not provide spaces for younger caregivers under 18 because of the liabilities minors produce. Having an online forum like Reddit or Quora specifically designated as having a target audience of dementia caregivers looking for helpful information and insights can change that dearth of support.

I've taken a look at real-world discussions on reddit, specifically of someone five years ago, who knows there is a need for a dementia caregiving resource guide. The post let everyone know it is for crowdsourcing that relevant caregiver support information. People share their lists and lists of notes they have taken, about everything spanning from Wyze cameras to "creating a spreadsheet for her life": bank accounts, medical info, IDs...the first response provides a plethora of links for all sorts of information, but I noticed that they were all in different webpages. Why not have one website that offers a comprehensive view? This post was 5 years ago, and just one example of caregiver discussions online. It shows the urgency to create something like this, and make it available and credible, preferably endorsed by global organizations like the Alzheimer's Disease International, which can allow caregivers to know they are not alone and have access to tips ("Resources"). While users currently use Reddit, AgingCare forums, the dispersed nature of these discussions shows the need for a centralized and reader-accessible database ("Caregiver Forum"; *ALZConnected*).

II. Inside and Outside Pools of Dementia Caregiving Within Families

Consider the differing stories of my dad's caregiving experience and my uncle's caregiving experience. According to my father,

- “We should always respect and cherish grandpa when we are caring for him. He’s taught us everything we know and guided us as we grew up. We should be eager to treat whatever his faults are through patience, always.”—Dad

On the other hand, my uncle recalls, “He can’t remember to eat or drink water. He forgets to wear his clothes and forgets about the directions back home. . . Dementia is so frustrating and it’s so annoying! It makes people so annoyed, in particular. Seeing grandpa, I do not wish to be that type of person when I’m a senior. I also don’t want these symptoms—there’s zero quality of life. There’s no joy and purpose in living anymore. It’s not worth it. Say anything off and he’ll hit you. He’ll interrupt other neighbors, but he doesn’t have that social awareness.”

As is evident, my dad believes a son should always express unwavering love for his ill dad. In contrast, my uncle dislikes everything about caring for his dad’s illness. These are the opposing sides of caring for the same person, and my dad and uncle will never get to see the other side because the amount of time they spend with my grandpa differs greatly.

My grandpa didn’t like having my uncle around. Sometimes, he hit him or was verbally aggressive. Some important context to understand is that my dad travels a lot, while my uncle rarely travels when he was taking care of grandpa. My dad doesn’t come back as often, perhaps two or three times a year, while my uncle is there everyday. This doesn’t mean that my dad does not contribute. He provides 90% of the money for my grandparent’s living expenses—including the pricey dementia costs: medication, hospital visits, the nannies, etc. All of it was prominently funded by my dad. My grandpa has a clear favorite child, and it is my dad, because he sees that he was an investment done well. My dad’s success thus brought him pride and joy.

When I asked my grandmother if there were any special happy moments she wanted to share after grandpa got dementia to see if not everything was anxiety-provoking, she initially said no, but then began to talk about how his face would light up whenever my dad was coming home to visit Dalian and see him. She recalled,

- “When your father comes home, he will be very excited. Your father, in his eyes, is probably the best son to him because he’s proud of his success. When I would say, he’s calling to talk to you, he’d be happy. When your dad bought gifts, clothes, mementos—sometimes I would use it as a strategy. When he would get mad or do something that needed changing, she would say I’m going to call your third son to tell you it’s wrong. He always thought whatever your dad said goes, he respected him the most, but everyone else he looked down on.”

This favoritism was reflected in the way he treated uncle versus my father. Hearing my father say to always tend to one’s parents tirelessly, it implied to me a lack of communication between my family members because of the significant discrepancy between my uncle’s and my father’s attitudes. My father didn’t understand what my uncle actually went through, seeing that he was gone two-thirds of each year. He didn’t quite have to worry about tiptoeing around my

grandpa to not make him mad like my uncle did, and I believe that if he had talked it through with my uncle about everything he told me in my interview with him, he would have a significant change of heart. Caregiving for dementia is more difficult than one may realize. One becomes the executive assistant responsible for ensuring all of a PWD's needs, and if the PWD lacks insight into their dementia, a PWD might incessantly argue with the caregiver for regulating their "perfectly healthy" life.

This juxtaposition is important in showing a real-world example of what miscommunication and misalignment of experiences can look like in a family. It emphasizes the importance of tackling discussions of family caregiving with sensitivity to the other person's unique experiences and the importance of proper communication. Telling what my father said to my uncle after an interview would be like disregarding all his feelings, saying that "well, still, no matter what you have to be complacent and happy to be caring for your father, who has done so much for you!" when you truly have not been there for the same difficulties. You cannot judge. It makes the other person feel guilty because they don't want to villainize a parent that has reciprocally cared for them their whole life, thus they feel singled out for their negative feelings about caregiving.

Society and current scholarly research regularly discuss the stigma that family caregivers face from people outside of family. But there's little discussion of how those same dynamics occur within families. Sometimes, opinions matter more when someone closer says it to you. In agreement with Smith et. al, several studies tend to solely focus on the relationship between the PWD and the family caregiver, and there's instead an uneven amount of work focused on the family network of the caregiver and how it influences their experience and process (Smith et al.). More work could be done in this area, and few studies also focus on the diversity of ways a family network as there are different categories of kinship that can affect the caregiving process as well. Understanding such family relationships can allow medical professionals and policy makers to treat and establish ways of support through informed advocacy. One notion from Smith et. al's work resonates with this paper: the idea that siblings can share the caregiving load to help alleviate the caregiving burden, but this sharing is commonly unbalanced, and there's an individual in the family who carries most duty in the end (Smith et al.). Moreover, communication difficulties can arise within family members as they experience an inability to understand or deny the struggles of the primary family caregiver (Smith et al.). This can all result in a receding community for the caregiver, where at this time of vulnerability, they may need the support and the presence of others willing to help out, leading to increased feelings of anxiety and worry.

Dementia needs to be discussed more in public and transparently about the hierarchies that form, specifically with those that take on different roles. A person who deals with monetary transactions is nevertheless essential, but may be less involved than a direct caregiver. Though my study is of a small sample size, I can hypothesize that the more proximal a caregiver is with a PWD, the more likely they are to experience relational conflict (Figure 1; Figure 2; Figure 3). Simultaneously, family support for a family caregiver can automatically shrink given that the

PWD can grow strongly negative feelings for a relative that reduces any sort of relationship the person once had with the PWD, causing them to visit less and engage less with supporting their dementia. Family members pull away because of their verbal attacks.

My cousin grew up around grandma and grandpa just like a son to them. He told me that he really loves them tremendously, cherishing their love more than even his paternal and maternal love. But he says that ever since my grandpa was diagnosed with dementia, he didn't understand social cues anymore and he wouldn't hold back his inner thoughts. He would say something hurtful, and then say, "I said it, so what?"

My cousin told me,

"Yes, I'm not that close to grandpa anymore. You know that he's ill and that you should love and care for him without a doubt, but in the end, some arguments hurt deeply. Specifically because there's a logical line of reasoning behind his insults, carrying an intent to hurt me. My dad and mom got divorced before he got dementia, and you know that they don't get along well. Hypothetically, by social etiquette, he shouldn't bring it up, but grandpa still connects those memories to me. When he wasn't ill, he didn't insult my mom. He probably held it back. But when he had dementia, he would say that "you did this just like your mom, you did it terribly" or "that wasn't good, that's what your mom would do." It's hard to respond optimistically to those statements, I can't help it. Especially because I prize my grandparents in my life, I felt especially defeated."

My cousin didn't visit my grandpa as frequently anymore, especially since his son wasn't welcomed. I asked my nephew to write me answers to some of my interview questions about his experience spending time around grandpa. He was around 5 or 6 when he visited the most and when my grandpa had dementia.

He wrote, "He didn't recognize me. But I don't blame him, because it's normal for old people to have a bad memory. He was both not good and good to me. I can't remember the good things, but the bad thing is that he had a bad temper. He would get really angry if we disturbed him or moved his things. He always spat everywhere, and I was afraid of stepping on it, so I didn't like going to his home". Finally when prompted by the question of what could be done to help his great-grandfather he said he wished he would talk to him, or help him calm down, so that he can feel "more comfortable in his heart."

As someone who also grew up alongside my grandpa's progressing dementia, I sometimes forgot that dementia is not normal. Therefore, I hypothesized that our aging population is slowly becoming more consolidated with the idea that dementia symptoms are normal and a regular component of the aging process because we are in a dementia epidemic. His responses suggest that growing up alongside a person progressing in dementia can shape the child's psychology when it comes to understanding the process of aging. For my nephew, he now has the impression that it's normal for old people to not recognize people. Memory is going to grow more inefficient as an individual gets older as the brain slows down somewhat, but not

recognizing someone at all is not normal. His experience reinforces my hypothesis that it is normalized as a standard feature of aging, when it is not.

Although my nephew's relationship grew increasingly remote with his great-grandpa, his experience is significant in that it is a representation of how perceptions of dementia ripple out across generations. Dementia is not just internal stress, it also shifts and distorts the broader cultural understanding of the illness of "outside pools".

Dementia can fracture family relationships. The previous examples reflect the isolation that my uncle experienced during caregiving, particularly from other family members who disregard the caregiving burden such as my dad, while my cousin and my nephew were pushed out of the inner circle and thus isolated by being forgotten entirely. This is what I mean by the title of this paper, his family tree recedes because certain branches are brought in (either welcomed or unwelcomed) while some branches are completely erased. The position of someone as they're experiencing dementia caregiving shapes the way they remember it.

While my father remembers caregiving as doing everything he can to make his father happy, buying him presents and just coming home seemed to do the trick and so he doesn't see the other side of the coin. Therefore, he wouldn't understand why someone might feel disrespected by his dad like my uncle and my cousin. In contrast, my uncle remembers caregiving as a time period of great mental strain, a packed schedule, and a constant need to be checking up on his parents, making sure they are fed, buying them groceries, ensuring the basic needs for survival.

My cousin remembers feeling targeted for his insecurity about his parent's divorce, insults to his mom made with no restraint, but always looking back at the tragedy that their healthy and close relationship unraveled this way. My nephew remembers disliking his visit back to that side of his family. He had empathy to help his grandpa but he did not know what to do because he didn't understand his condition. He attributed it to a normal trait that older people assume.

This suggests that support systems should not be limited to only targeting one prominent family member. Rather, they should be oriented to the entire family network, detangling social hierarchies and smoothing out communication and providing a space to share their specific experiences so that everyone can understand each other better, work towards a better system, distribute work and responsibility more evenly if they have more awareness of the struggles of each person. They should also be implemented in policy, in initiatives where counselling for the whole family is integrated into a medical insurance program. There could be some forms of publicly funded advisors, in hopes that a subsidized cost for such resources that are usually only limited to affluent populations to the ability to allocate more money—can also reach more rural and lower-income communities. Those advisors can help alongside the family to put together a caregiving plan, and mediate conflicts they bring in.

For the younger ones growing up beside an elder with dementia or even young-onset dementia in their parents, it would be helpful to incorporate some information about dementia

and other life-changing illnesses, like cancer, into school curriculum or proceed with workshops in health centres to further educate and cause less misconception.

Discussion of AI in Dementia Caregiving

In global research and invention addressing the rising issue of dementia, applications of artificial intelligence (AI) have caught the heat of contemporary discussion. As an illness with no current cure, AI seems like the next best thing other than the official creation of a disease-modifying treatment. AI holds promise in mediating all sorts of sectors in caregiving: diagnostics, companionship, social interaction, monitoring, cognitive training, are just a few of the possibilities that can be further explored and have already been tapped into by companies (Ma et al.). While the innovation and pursuit of new AI technologies requires considerable financial resources and comes with the duty of monitoring the technology itself for possible machine learning mistakes, AI can greatly reduce a quality sometimes ineluctable with human caregiving: human error. The night my grandpa passed away was caused by our nanny's mistake. As our nanny, she accepted the responsibility to wake up every night to guide my grandpa when he needed to use the bathroom. Since my grandpa had impaired visuospatial abilities and worsened coordination, this was a step absolutely necessary to take, and we needed someone to monitor him. One night, my grandpa needed to go to the washroom, but our nanny did not wake up. Soon after, when my grandpa was trying to find his way to the bathroom, he bumped into a spatial barrier and fell to the ground. Our nanny only discovered him the coming morning, but he had broken his spine and it was too late.

There's no closure in blaming our nanny, but this is a direct example of how human error can cause tragedies. When it comes to someone's life, you want to limit the possibility of this happening to the best of your ability. AI can indeed help us reach that goal. According to Qi et al, several countries have seen benefits arise from implementing telepresence robots to improve home intervention with dementia patients. Studies show that the American Socially Assistive Robots (SARs) presents to users "safety monitoring and seeking emergency assistance", which would have been incredibly useful for my grandpa that night, but also for the apparent 80% of dementia patients that fall annually as well (Qi et al.; "Meta-Analysis-Based Comparison"). Alzheimer's Research UK reported in 2024 that people with dementia are three times more likely to get seriously injured after a fall and end up admitted to the hospital (Taylor). People with dementia also have a higher probability of having surgical site bleeding and sepsis compared to people who don't have dementia. Both of these conditions are associated with higher death rates (Taylor).

In a study by Wu et. al, covering the responses and feedback of formal healthcare professionals (HCP) on the SAR "Archie", NAO robot from SoftBank Robotics, they found that participants took note of the positive benefits of such technology. Still, in order to facilitate effective adoption of SARs into where they are needed in healthcare, participants emphasized the importance of bettering the technology's usability and that different strategies should be used depending on the stage of dementia (Wu et al.). One HCP criticism of Archie was that, "someone

who exhibit[s] more moderate symptoms of dementia may be really scared by the noise, the lights, the movement” (Wu et al.). Similarly, another HCP brought up how “robot anxiety” may present a barrier to the technology’s acceptance in PWDs. However, a third HCP said that these challenges might be “compensated for by just introducing Archie earlier into their progression” (Wu et al.).

Nevertheless, SARs technology is still a work in progress in the dementia domain, but there’s also a lot of momentum surrounding the subject and the tool’s advancement. This progress and the research about it serves as a representative of one of the complexities in AI caregiving intervention, but it also shows there is hope that through greater research, we can find better ways to alleviate the stress on a caregiver’s shoulders all while making the dementia patient’s experience better.

Conclusion

Dementia caregiving is a multi-layered and mutable practice, shaped by systemic, cultural, and sentimental pillars including caretaker burnout, relational conflict, coping strategies, filial piety, seniority, support, and affection. It is a topic in research that cannot be fully understood by extrapolating from just one case study. The more research we have, using qualitative and quantitative methods, can help us understand a comprehensive view that can bring to surface ways to strategize, mitigating the impact the journey of dementia progression has on the primary family caregiver and the entire network. I proposed how establishing a composed and concentrated dementia caregiving database can be of utility to caregivers looking for answers and tips. Life story work should be used at every advantage. It can provide caregivers a technique in which PWDs can share their stories and identities in an accessible way to people who need to care for them in a more personalized way. I have discussed how whole family-oriented counselling and advising help alleviate caregiver stress, as well as the importance of conducting workshops or short educational modules for youth who need clarification on their knowledge of dementia. Finally, I have referred to the rise of AI technology pursuits in our dementia epidemic as promising and grounded in a true need for the tools. My family’s experience with my grandpa has guided my research and sustained my ambitions to better understand what caregiving narratives can look like across generations, showing me that as we move closer to understanding the effects dementia has on the entire family, we can implement more specialized and pinpoint strategies that improve policy, education, and healthcare, one by one, changing the experiences of dementia caregiving forever, hopefully to allay future generations’ fears and concerns.

Acknowledgements

I’d like to thank Professor Angela Garcia, at Stanford’s Anthropology Department, for mentoring me and supporting this endeavour while I explored this rich and fascinating subject of study. I’d also like to thank my grandpa for bringing me an intimate personal recollection to offer in the greater discussion of dementia caregiving.

Works Cited

- ALZConnected. alzconnected.org/discussions. Accessed 26 Aug. 2025.
- Alzheimer's Disease International, et al. World Alzheimer Report 2024. Alzheimer's Disease International, 20 Sept. 2024. Alzheimer's Disease International, www.alzint.org/u/World-Alzheimer-Report-2024.pdf. Accessed 14 Aug. 2025.
- "Anterograde Amnesia." Yale Medicine, www.yalemedicine.org/clinical-keywords/anterograde-amnesia. Accessed 26 Aug. 2025.
- "Caregiver Forum." AgingCare, www.agingcare.com/caregiver-forum. Accessed 26 Aug. 2025.
- Chainay, Hanna, et al. "Emotional Enhancement of Immediate Memory: Positive Pictorial Stimuli Are Better Recognized than Neutral or Negative Pictorial Stimuli." *Advances in Cognitive Psychology*, vol. 8, no. 3, 21 Aug. 2012, p. 256. National Library of Medicine, [pmc.ncbi.nlm.nih.gov/articles/PMC3434683/#:~:text=The%20idea%20is%20that%20the, and%20consequently%20improve%20subsequent%20recognition](https://pubmed.ncbi.nlm.nih.gov/articles/PMC3434683/#:~:text=The%20idea%20is%20that%20the, and%20consequently%20improve%20subsequent%20recognition). Accessed 25 Aug. 2025.
- Che, Hui-Lian, et al. "The Self-Empowerment Process of Primary Caregivers: A Study of Caring for Elderly with Dementia." *Journal of Nursing Research*, vol. 14, no. 3, Sept. 2006, pp. 209-18, <https://doi.org/10.1097/01.jnr.0000387579.58974.0b>.
- "Creating a Life Story for a Person with Dementia." DementiaUK, Oct. 2023, www.dementiauk.org/information-and-support/living-with-dementia/creating-a-life-story/. Accessed 26 Aug. 2025.
- "Dementia Statistics." Alzheimer's Disease International, www.alzint.org/about/dementia-facts-figures/dementia-statistics/. Accessed 14 Aug. 2025.
- England, Paula. "Emerging Theories of Care Work." *Annual Review of Sociology*, vol. 31, no. 1, 1 Aug. 2005, pp. 381-99, <https://doi.org/10.1146/annurev.soc.31.041304.122317>.
- "Filial Piety." ScienceDirect, Elsevier, www.sciencedirect.com/topics/psychology/filial-piety. Accessed 25 Aug. 2025.
- Irish, Muireann. "Autobiographical Memory in Dementia Syndromes—An Integrative Review." *WIREs Cognitive Science*, vol. 14, no. 3, 14 Oct. 2022, <https://doi.org/10.1002/wcs.1630>.
- Kindell, Jacqueline, et al. "Life Story Resources in Dementia Care: A Review." *Quality in Ageing and Older Adults*, vol. 15, no. 3, 2 Sept. 2014, pp. 151-61. National Library of Medicine, <https://doi.org/10.1108/qaoa-02-2014-0003>.
- . "Everyday Conversation in Dementia: A Review of the Literature to Inform Research and Practice." *International Journal of Language & Communication Disorders*, vol. 52, no. 4, 27 Nov. 2016, pp. 392-406, <https://doi.org/10.1111/1460-6984.12298>.
- Kintz, Stephen, and Heather Harris Wright. "Semantic Knowledge Use in Discourse: Influence of Age." *Discourse Processes*, vol. 54, no. 8, 29 Feb. 2016, pp. 670-81. National Library of Medicine, <https://doi.org/10.1080/0163853x.2016.1150652>.

- Knecht, Hanna L., and Francisca S. Rodriguez. "Social Interactions in Dementia: Perceptions of Current Situation and Opportunities." *BMC Geriatrics*, vol. 25, no. 1, 27 Mar. 2025, <https://doi.org/10.1186/s12877-025-05850-8>.
- Ma, Bingxin, et al. "Artificial Intelligence in Elderly Healthcare: A Scoping Review." *Ageing Research Reviews*, vol. 83, Jan. 2023, p. 101808. ScienceDirect, <https://doi.org/10.1016/j.arr.2022.101808>.
- "Meta-Analysis-Based Comparison of Annual Fall Risk between Older Adults with Alzheimer's Disease and Mild Cognitive Impairment." *Advances in Geriatric Medicine and Research*, 2024. National Library of Medicine, <https://doi.org/10.20900/agmr20240002>.
- "Multiple Trace Theory." Taylor and Francis, taylorandfrancis.com/knowledge/Medicine_and_healthcare/Neurology/Multiple_trace_theory/. Accessed 25 Aug. 2025.
- Nadel, L., et al. "Multiple Trace Theory of Human Memory: Computational, Neuroimaging, and Neuropsychological Results." *Hippocampus*, vol. 10, no. 4, 2000, pp. 352-68. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/10985275/.
- Nichols, Katherine R., et al. "When Dementia Is in the House: Needs Assessment Survey for Young Caregivers." *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*, vol. 40, no. 1, Jan. 2013, pp. 21-28. National Library of Medicine, <https://doi.org/10.1017/s0317167100012907>.
- Qi, Jinghan, et al. "Artificial Intelligence (AI) for Home Support Interventions in Dementia: A Scoping Review Protocol." *BMJ Open*, vol. 12, no. 9, Sept. 2022, p. e062604. National Library of Medicine, <https://doi.org/10.1136/bmjopen-2022-062604>. Accessed 27 Aug. 2025.
- Reed, Jonathan M., and Larry R. Squire. "Retrograde Amnesia for Facts and Events: Findings from Four New Cases." *The Journal of Neuroscience*, vol. 18, no. 10, 15 May 1998, pp. 3943-54. National Library of Medicine, <https://doi.org/10.1523/jneurosci.18-10-03943.1998>.
- "Resources." Reddit, 2020, www.reddit.com/r/dementia/comments/l750sg/resources/. Accessed 26 Aug. 2025.
- Smith, Lourett, et al. "Family Dynamics in Dementia Care: A Phenomenological Exploration of the Experiences of Family Caregivers of Relatives with Dementia." *Journal of Psychiatric and Mental Health Nursing*, vol. 29, no. 6, 2 Feb. 2022, pp. 861-72. Wiley Online Library, <https://doi.org/10.1111/jpm.12822>.
- Subramaniam, Ponnusamy, et al. "Life Story Book to Enhance Communication in Persons with Dementia: A Systematic Review of Reviews." *PLOS ONE*, vol. 18, no. 10, 5 Oct. 2023, p. e0291620. National Library of Medicine, <https://doi.org/10.1371/journal.pone.0291620>.
- Taylor, Emma. "How Does Dementia Cause Death?" *Alzheimer's Research UK*, 9 Dec. 2024, www.alzheimersresearchuk.org/news/how-does-dementia-cause-death/. Accessed 27 Aug. 2025.

- Wang, A.-Y, et al. "Socioeconomic Status and Risks of Cognitive Impairment and Dementia: A Systematic Review and Meta-Analysis of 39 Prospective Studies." *The Journal of Prevention of Alzheimer's Disease*, vol. 10, no. 1, Jan. 2023, pp. 83-94. ScienceDirect, <https://doi.org/10.14283/jpad.2022.81>.
- Wu, Xian, et al. "Socially Assistive Robots for Dementia Care: Exploring Caregiver Perceptions of Use Cases and Acceptance." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, no. 1, Sept. 2021, pp. 6-10. National Library of Medicine, <https://doi.org/10.1177/1071181321651192>.

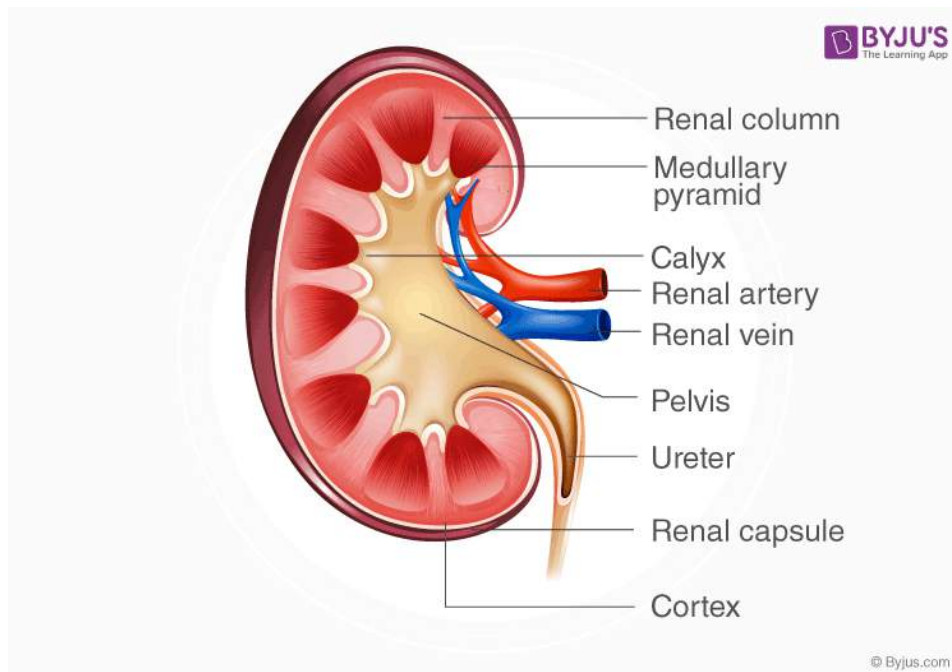
Renal Cell Carcinoma: Epidemiology, Risk Factors, Genetics, and Treatment.

By Sabrina Kainbacher

Abstract

Renal Cell Carcinoma (RCC) is the most common type of kidney cancer. It has affected the world with a mortality rate of 30 to 40%, which is more commonly seen in men than women (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). With improvements in diagnostic tests, such as CT and MRI imaging, a drastic amount of diagnosed patients has grown rapidly in RCC. There are many risk factors which include obesity, hypertension, smoking, and chronic kidney disease. The most common type of RCC is clear cell carcinoma that frequently occurs due to mutations in the *VHL* gene. The most selective technique to maintain kidneys in patients is Nephron Sparing surgery, but surgeries to remove small tumors include radical nephrectomy and partial nephrectomy. Patients with metastatic RCC are treated with surgical approaches, adjuvant therapy, and targeted therapy. This article covers RCC including its epidemiology, risk factors, pathophysiology, genetics, and treatments.

Introduction



The image above shows a healthy kidney with the basic parts.

Kidney cancer starts in the kidney, which are two bean-shaped organs. They are located behind the abdominal organs with one kidney on each side of the spine. The kidney's purpose is to expel excess water, salt, and waste products from your body and turn those substances into urine. Kidneys are made up of cellular tissue, but sometimes those tissues behave irregularly.

When there are changes in a kidney's DNA, they start to grow in an abnormal way and form tumors, which is how kidney cancer develops. RCC forms in the tubules, which are tiny tubes inside a kidney. The tubules' purpose is to directly move substances a person's body needs. For example, water and nutrients go to a person's bloodstream while filtering waste through your urine. The kidney's symptoms include: pain, blood in urine, hematuria, and palpable abdominal mass that occurs in 4-17% of cases.

Renal Cell Carcinoma is the most common kidney cancer there is. It has affected the world greatly with the cause of 3% of malignancies in women, and the tenth most common cancer in women (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). In men, RCC is the sixth most common kidney cancer, and it is the cause of 5% of malignancies (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Malignancies are when abnormal cells divide without control and can invade nearby tissues. Additionally, the most common kidney cancer for adults is Renal Cell Carcinoma (RCC). It mainly occurs in Europe and North American populations, but according to the Global Cancer Statistics in 2020, the amount of RCC-related deaths were 179,368 and 431,288 people were diagnosed with RCC (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The death rate of RCC in 2016 has concluded to approximately 2% of all cancers (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The diagnoses have been significantly growing, leading to improved cross-section abdominal imaging. Even though RCC is the most common urogenital malignancy, it is commonly misdiagnosed. Optimal screening modalities have not been established yet, which is critical in reducing the amount of deaths in RCC. Therefore, having patients diagnosed in early stages is crucial in treating patients and reducing the amount of deaths.

Epidemiology

RCC is a very complex and heterogeneous disease with several clinical features. For example, RCC was recognized as the seventh most cause of cancer with over 140,000 deaths in 2013 (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). RCC is considered 2-3% of all cancers (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). RCC diagnoses are continuously increasing today, which varies across the world and is higher in developed countries than developing countries. RCC mainly occurs at the ages of 60 to 70 years old, but declines after 70 years old, which is due to less use of aggressive diagnostic testing within the age range (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The increase in the diagnosis of RCC is the improvement of diagnostic procedures and public consciousness of the importance of periodic health screenings. As a result, the amount of patients with RCC being diagnosed at early stages has drastically increased. Furthermore, due to early diagnosis and therapy, the mortality rate has dropped, but metastasis, the process where cancer cells spread to other parts of the body, still occurs in many patients.

However, the change has still contributed to improvements, with 20-30% of patients being metastatic at an early stage of diagnosis (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Secondly, 30-50% of patients are progressing to metastasis. Lastly, 40% of patients with localized RCC have distant metastases after surgery (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The lungs, bones, opposite kidney, and liver are all common distant metastases.

Risk Factor

In RCC, the main major risk factors are age and gender. Other risk factors include location, ethnicity, history of smoking, history of using tobacco products, obesity, and hypertension. Multiple risk factors which may have an association with RCC include chronic kidney disease, acquired renal cystic disease, end stage renal disease, chronic use of palliatives, exposure to cadmium and trichloroethylene, consumption of red and processed meat, coral hepatitis, vitamin D level, type-2 diabetes, increased triglycerides, decreased physical activity, and genetic syndrome.

Genetics

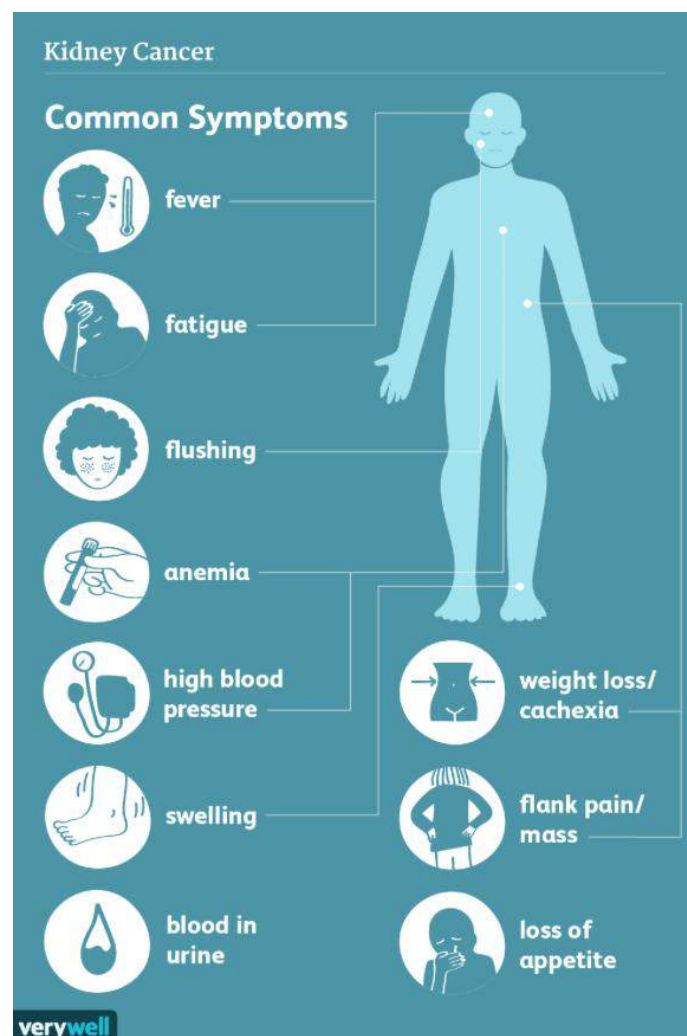
WHO classification of renal cell carcinoma (2016)
Clear cell renal cell carcinoma
Multilocular cystic renal neoplasm of low malignant potential
Papillary renal cell carcinoma
Hereditary leiomyomatosis and renal carcinoma-associated renal cell carcinoma
Chromophobe renal cell carcinoma
Collecting duct carcinoma
Renal medullary carcinoma
MiT family translocation renal cell carcinomas
Succinate dehydrogenase-deficient renal carcinoma
Mucinous tubular and spindle cell carcinoma
Tubulocystic renal cell carcinoma
Acquired cystic disease-associated renal cell carcinoma
Clear cell papillary renal cell carcinoma
Renal cell carcinoma, unclassified
Papillary adenoma
Oncocytoma

The image above shows the WHO classification by Heidelberg (2016).

RCC is a type of cancer that has many genetic and epigenetic alterations. The first classification of RCC was by Heidelberg in 1997. She performed the classification on molecular genetics, which was entered as the WHO tumor classification in 2004, 2012 Vancouver ISUP, and the classification of the WHO in 2016 (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). About 3% of the RCC cases have a similar background with an autosomal predominant pattern, which showed that RCC can be divided into heredity types (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). This is uncommon. Subtypes of RCC have multiple different mutations and epigenetic alterations in the genes that cause RCC. The important genes

that play an important role are the *VHL*, *MET*, *FH*, *BHD*, and *HRPT2* (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The mutation which is most common is the gene *VHL* which causes the heredity of clear cell RCC. In clear cell RCC, *VHL* is the tumor suppressor which is seen in patients with susceptibility to von HippelLindau inherited syndrome. But this type of genetic alteration has the best prognosis among patients with clear cell RCC. 3p25-26, 3p12,3p14.2, 3p21.1, 3p21.3, 3p22, and 3p26.2 all are tumor suppressor genes, which could be responsible for the initiation progression or inactivation of those genes in chromosome 3 (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The gene *PBRM1*, which involves tumor growth and metastasis, is a potential gateway for predicting the prognosis of clear cell RCC.

Signs and symptoms



The image above shows the different symptoms of RCC

Random detection of renal masses has significantly increased in following routine imagery for various medical disorders. So, more than 50% of patients with RCC are being

diagnosed accidentally. A renal mass could just be a simple kidney cyst that could have no treatment involved, but in some cases it isn't (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Some benign renal lesions or malignant RCC require interventions. About 30% of patients with RCC are diagnosed with symptoms, but small masses usually do not cause any symptoms (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). This leads to an increase in growth and build up over time, resulting in metastatic RCC at the time of diagnosis. The symptoms include having abdominal pain, fever, hematuria, weight loss, anemia-induced fatigue, or secondary symptoms caused by metastasis spread such as bone pain and cough. When all stages are completed, RCC may produce biological pseudo-hormonal or pseudo cytokine products, leading to clinical paraneoplastic syndromes. This includes: hypertension, anemia, cachexia, weight loss, fever, polycythemia, hypoglycemia, hypercalcemia, liver function disorders, and neuropathy. Some signs in physical examination may have a limited role in diagnosing RCC, but some signs may be important such as abdominal mass, peripheral lymphadenopathy, and lower limb edema.

Treatment

Nephron-sparing surgery

One treatment is surgery, specifically nephron preservation surgery. Nephron sparing radiation nephrectomy is the removal of the entire kidney, but when using the nephron-sparing surgery technique, only the tumor is removed. Moreover, during the technique the normal kidney parenchyma is preserved but the negative margins of surgery are sufficient. When it is possible to maintain part of the kidney, using the nephron-sparing treatment is very selective for the job. The treatment is very efficient with the same oncological control and the additional benefit for maintaining the kidney. It is also less risk of postoperative cardiovascular occurrences compared to radical nephrectomy. For T1 tumors, partial nephrectomy should be used, but if the tumor is in a location that is not possible to be removed, partial nephrectomy won't be an option. For T2 tumors, partial nephrectomy is also an option, but it depends on the location as well. By the guidelines of European Association of Urology, the standard indications for nephron-sparing surgery include: 1) Absolute: Applied for patients with an anatomical kidney or a functional kidney. 2) Relative: Applied for patients with functional kidney, a disorder that may damage the kidney function in the future or hereditary form of RCC that has a high risk of developing tumors in the kidney. 3) Elective: Applied for patients with one lateral localized RCC with a healthy kidney. Nephron sparing surgery is mainly used for metastasis and locally advanced diseases.

Nephron sparing surgery is not commonly used for RCC since the techniques require experience and may be associated with a risk of bleeding and urinary leakage. When using nephron sparing surgery instead of radical nephrectomy, it depends on the diameter, location, depth, and proximity to the hilar vessels and urinary collecting system. There are four best scores which are used to evaluate if a patient should have surgery or not. The scores are aimed at

improving patient selection, surgical management, research report, and outcome prediction including C-index, RENAL, surgical approach in nephron-sparing surgery. This depends on the experience of the surgeon and the tumor's characteristics. The approaches are similar to the terms of surveillance. There are 16-35% of hospitals that utilize robotic surgery systems for nephron-sparing treatment (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022).

Laparoscopic partial nephrectomy

A second type of surgery for RCC is laparoscopic partial nephrectomy, which takes longer than nephron sparing surgery. In laparoscopic partial nephrectomy, there is a higher risk of kidney ischemia and complications are higher, which is why partial nephrectomy is preferred. Laparoscopic nephron-sparing surgery is only used if the surgeon has more experience. In the 2000s, open nephron-sparing surgery was the standard treatment, but now it is only used for cases of anatomical difficulties like large tumors or urinary tract infections, or for patients with low invasive techniques (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Radical nephrectomy and partial nephrectomy should be used for smaller tumors. Radical nephrectomy is useful for T2 tumors and larger tumors. Adding on to that, radical laparoscopic nephrectomy is the standard treatment for T2 tumors for patients. Compared to open radical nephrectomy, there is less blood during surgery, less need of painkillers, and a shorter recovery. Radical open nephrectomy has upsides too though. It is a standard surgical procedure for T3 and T4 tumors.

Adjuvant therapies

Although surgery is used mostly for RCC, 20-30% of patients have distant metastasis and 2-5% of patients have local recurrence (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Due to this, adjuvant therapy is important after surgery. The approaches are hormone therapy, radiotherapy, immunotherapy, vaccine, and target agents.

Immunotherapy treatment for metastasis disease

Immunotherapy for metastasis RCC only does a complete response in less than 10% of patients and the side effects are severe with toxic effects in the body (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The only treatments that are safe include modulating drugs that are approved are Aldesleuki, INFa, and Bevacizumab. INFa is used in schedules, doses, and multiple formulas with a response rate of 10-15% (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The average response rate is four months and there is no relationship between the dose and the response (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Three times a week equals a dose of 5-10 million units per body surface area (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., &

Mafakher, L. 2022). Even though INF has good performance status, it is not recommended to be treated as a single drug because IL2 causes tumor regression by activating the immune system in RCC. During the 1990s, treatments with a high dose of IL2 were commonly used for patients with good PS and organ function. IL2 has various severe toxicity: severe hypertension, fluid retention, vascular leak syndrome, multiple organ failure, *axotomy*, oliguria, cardiac arrhythmia, fever, metabolic acidosis, dyspnea, and skin complications are all common side effects (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). Patients need severe care with a mortality rate of 2-4%, and due to the toxins in high doses of IL2, studies are encouraging other doses and methods (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). The morbidity and mortality rate is important in high dose IL2 selected patients causing long full-term implantation. High dose IL2 is still an option for young patients in very good conditions and low tumor volume, including patients with isolated pulmonary metastases in experienced centers.

Other treatment options for metastasis disease

The most common options for patients with metastasis RCC is with the targeted therapies including Ca, corrected HB serum, KPS, neutrophil count, and platelet count (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). It seems through the use of radiotherapy, part of a multimodality approach, RCC tends to be radio-resistance tumor. This is because of its poor results with low-dose RT.

Conclusion

Renal Cell Carcinoma is the most common type of kidney cancer with a mortality rate of 30%-40% (Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. 2022). In most cases, RCC is diagnosed accidentally. The diagnoses of patients with RCC has drastically increased due to the improvements in diagnostic tests. Having patients with RCC diagnosed at early stages is critical in treating patients and reducing death rates. Therefore, optimal screening modalities and approaches have not been established yet. Having knowledge of the recent advances in diagnosis and management of RCC could help physicians and nephrologists to better treat and diagnose Renal Cell Carcinoma.

Works Cited

- Chengyu You, Yuelin Du, Hui Wang, Lei Peng, Tangquiang Wei, Xiaojun Zhang, Xianhui Li, Anguo Wang. (28 October 2020). Laparoscopic Versus Open Partial Nephrectomy: A Systematic Review and Meta-Analysis Of Surgical, Oncological, and Functional Outcomes. US: Frontiers. Retrieved from <https://doi.org/10.3389/fonc.2020.583979>
- Bahadoram, S., Davoodi, M., Hassanzadeh, S., Bahadoram, M., Barahman, M., & Mafakher, L. (2022). Renal cell carcinoma: an overview of the epidemiology, diagnosis, and treatment. *G Ital Nefrol*, 39(3), 2022.
- Cleveland Clinic. (2023). Renal Cell Carcinoma. US: Cleveland Clinic. Retrieved from <https://my.clevelandclinic.org/health/diseases/24906-renal-cell-carcinoma>

How BRICS CBDC can Reduce Dollar Dominance By Tariq Murgian

Over the past several years, the development of central bank digital currencies (CBDCs) has gained increasing attention as a tool for reshaping global currency dynamics. Literature reviewed in this paper highlights how CBDCs can reduce reliance on dominant currencies, while intra bloc coordination amplifies their potential impact. Analysis of BRICS economies demonstrates that a shared CBDC could influence three key channels: international trade, financial services, and remittances. Among these, trade appears most immediately effective, while financial services and remittances offer structural and social support for broader adoption.

Keywords BRICS CBDC, Dollar dominance, remittances, international trade, financial services

Section 1: Introduction

To understand the role of a BRICS digital currency in financial services, it is important to look at how cross border settlement currently works within the bloc. A BRICS digital currency has transformative potential for the cross border financial services landscape within the bloc. Currently, even when financial flows such as loans, portfolio investments, interbank transfers, and corporate transactions occur between BRICS entities, the U.S dollar remains the principal currency of settlement. As seen in Graph 3, the dollar still makes up the majority of global reserves, though its share has been gradually declining, while the renminbi remains small but is steadily increasing. This reliance persists despite both counterparties using local BRICS currencies domestically, and it is sustained by the deep liquidity and historical dominance of the dollar in global finance.

What makes the BRICS bloc distinct is the scale and complementarity of its economies. Together, they represent over 30% of global GDP and an unusually diverse mix of exporters, importers, and commodity producers. Unlike regional blocs such as ASEAN or the EU, BRICS operates across multiple continents and political systems, giving it a broader platform to test a shared digital mechanism without being bound by regulation. This diversity, often seen as a weakness, also provides resilience, as shocks in one member's currency can be offset by stability in another.

The BRICS countries Brazil, Russia, India, China, and South Africa are each exploring CBDCs in different ways. CBDCs matter in this discussion because of their potential impact on international trade and currency dependence. If BRICS countries can use CBDCs to settle transactions with each other without converting to dollars, it could reduce the global demand for dollars in cross border trade. Currently, however, the U.S dollar still accounts for the majority of global trade settlements and official foreign exchange reserves, underscoring its dominant role in the system. This paper thus sets out to answer the question of how a BRICS CBDC can reduce reliance on the dollar and reshape financial flows within and beyond the bloc.

While this paper explores the potential of such a shift, it also recognizes that no digital currency, BRICS or otherwise, is immune to challenges. The goal here is not to predict the

complete displacement of the dollar, but to evaluate whether a BRICS CBDC could reduce some of the structural imbalances that the dollar system currently reinforces, and to assess under what conditions that transition could succeed.

Current research has shown that CBDCs are studied across three main branches of literature. The first focuses on monetary policy and financial stability, highlighting how digital currencies could alter transmission mechanisms and impact bank intermediation (Bindseil et al., 2020). The second examines cross border payments, noting that CBDCs could lower costs, increase speed, and reduce reliance on dominant currencies like the U.S dollar (BIS et al., 2021). The third branch looks at geopolitical and international trade implications, where scholars argue that initiatives such as a BRICS CBDC could strengthen regional integration while challenging dollar dominance (Subacchi et al., 2021). Together, this literature suggests that while risks remain, CBDCs are increasingly seen as a tool not only for domestic modernization but also for reshaping global currency dynamics.

A BRICS digital currency could reshape three major areas of economic activity: trade, financial services, and remittances. In trade, most BRICS countries still invoice transactions in dollars, which exposes exporters and importers to exchange rate volatility and adds conversion costs even when both sides use local currencies domestically. In financial services, even interbank transfers and corporate transactions between BRICS members are typically settled in dollars, reflecting its 58% share of global foreign reserves, though the renminbi is beginning to rise. A shift toward a shared digital currency would reduce reliance on the dollar, lower settlement costs, and gradually build liquidity within the bloc's own financial system. The remittance sector adds a third dimension: India alone is expected to receive over \$120 billion in remittances by 2025, and current systems require costly conversions across multiple BRICS currencies. A unified CBDC could streamline these flows, reduce fees, and provide transparency, offering both economic efficiency and greater financial inclusion.

Comparing the outcomes across the cases examined, it is clear that countries with more integrated financial systems experienced less volatility in trade flows and more stable investment patterns. In contrast, nations with fragmented systems faced higher disruptions, particularly in sectors dependent on cross border financing.

These differences highlight the importance of targeted policy interventions to strengthen financial connectivity and resilience. Implementing frameworks that enhance transparency and cooperation can mitigate shocks and support sustainable economic growth.

Section 2: Literature Review

1) Currency dominance

There have been many papers that studied currency dominance, and in particular the dominance of the U.S dollar (USD) in global finance and trade. Gopinath et al. (2020) showed that there is strong evidence of U.S. dollar dominance in international trade. They found that the USD is used to price most international trade transactions, even when the United States is not

one of the countries involved. Their study used trade data from over 2,500 country pairs covering approximately 91% of global trade and firm level data from Colombia to demonstrate how dominant the dollar is in price setting. This paper is important in showing how trade is often not priced in local currencies but in a few “dominant” ones, especially the dollar.

Gopinath et al. (2020) also explained how this pattern reinforces itself: once a majority of firms begin pricing in the dollar, it becomes difficult for others to move away from it. Their methodology combined macro level and micro level analysis, showing not only global trends but also how firms in emerging economies react to exchange rate movements. Their findings are highly relevant to this paper, as they suggest that for a BRICS central bank digital currency (CBDC) to challenge the dollar, it would need to become widely used in invoicing and accepted across trade networks. Researchers have shown that the U.S dollar became dominant because of its widespread usage, long standing stability, and central role in financial markets (Gopinath et al., 2020; Eren and Malamud, 2019). Its use spans across global trade, reserves, and payments, while its stability backed by U.S institutions and consistent monetary policy makes it a reliable store of value, especially during periods of economic uncertainty.

Building on its dominance in trade, the USD has also become the leading currency in global borrowing and financial markets. Eren and Malamud (2019) argued that the dollar is the most commonly used currency for issuing international debt, particularly because of its trusted value and the depth of U.S capital markets. They found that companies prefer borrowing in USD especially during periods of crisis because U.S monetary policy tends to respond to downturns with interest rate cuts. As a result, debt held in USD becomes cheaper to repay when global conditions worsen. Their methodology included a theoretical model of international debt issuance combined with real world data from global bond markets between 2000 and 2019. Although their focus was on the crisis period, the results imply that the USD is attractive in both stable and unstable times.

This suggests that a BRICS CBDC must replicate some of the benefits of the dollar such as low cost borrowing, policy predictability, and trust in the system if it wants to compete globally. In other words, it's not enough to introduce a new currency; it must also provide the financial advantages that make the dollar appealing in the first place.

In a broader economic sense, Pettis (2023) studied the long term effects of U.S dollar dominance on the American economy. He argued that being the world's reserve currency forces the U.S to run persistent trade deficits so that the rest of the world can access enough dollars. This results in negative side effects like industrial decline, increased debt, and higher unemployment in the U.S. BRICS members are affected by this dynamic because their export-led growth relies on a system that channels global demand toward the U.S economy, limiting their own domestic development and creating external vulnerabilities. Pettis (2023) took a historical and theoretical approach, analyzing trends from the post World War II period to the present. His findings help explain why some countries including members of BRICS would want to reduce reliance on the dollar and push for their own global currency instead.

Another key paper that supports this view is by Coquidé, Lages, and Shepelyansky (2023). Using trade data from 194 countries and network based simulation models, they found that countries with closer trade ties to BRICS members were more likely to shift toward using a BRICS led currency. Their paper, though focused on a hypothetical “BRI” currency, shows how currency dominance can shift over time through changes in trade relationships.

In addition to trade and finance, the U.S dollar plays a significant role in global remittances. According to the World Bank (2023), over 60% of global remittance flows are either denominated in or routed through the USD due to its central role in international banking systems. This is because most international banks hold dollar reserves, and the USD is the default currency for settling payments across borders. Global payment systems are built around U.S based financial institutions, making it easier and more efficient to process transfers in dollars. This highlights how dollar dependence extends beyond governments and corporations to individuals especially in developing countries. A BRICS CBDC that offers faster, cheaper, and direct transfers could provide an alternative and reduce reliance on the USD for millions of workers sending money home.

In sum, the literature on currency dominance highlights that the USD became dominant through a combination of widespread trade invoicing, financial trust, and policy reliability. However, it also shows that this dominance comes with costs for both the U.S. and other countries and that shifting away from it, while difficult, is not impossible. For a BRICS CBDC to be successful, it must not only be technologically functional but also offer the economic stability and network trust that underpin dollar dominance today.

2) Development of CBDC's

In recent years, the development of Central Bank Digital Currencies (CBDCs) has gained global momentum, with over 130 countries now exploring or piloting some form of digital currency (IMF, 2023). While some countries like the Bahamas, Nigeria, and Jamaica have already launched retail CBDCs, these are meant for public use, unlike wholesale CBDCs which are reserved for financial institutions. Others such as China, India, and Brazil are in advanced stages. Wealthier countries including the U.S and members of the EU remain cautious, focusing more on innovation, privacy, and competing with technology. (Nottingham Trent University, 2024).

Additionally, recent research has found that the adoption of CBDCs is not only shaped by technological or economic readiness but also by political factors. According to a study by Nottingham Trent University (2024), more authoritarian governments are leading the charge on CBDCs. Their centralized power structures make it easier to implement tightly controlled systems with little public debate. These governments often aim to increase control over payments and gain visibility into financial flows. CBDCs rely on centralized power structures, which make it easier for governments to implement tightly controlled systems. In many cases, governments pursue CBDCs to increase oversight of payment systems and gain visibility into financial flows.

For developing countries the appeal is more practical therefore improving financial inclusion and reducing corruption.

However, developing a CBDC comes with significant challenges especially for developing economies. Nugraha (2024) identified six major barriers: limited internet access, low digital and financial literacy, lack of trust in government, privacy concerns, weak legal frameworks, and little incentive for users to switch from cash or mobile money. These issues are especially important for BRICS countries like South Africa or India, where rural populations may struggle with digital access. For any future BRICS CBDC to succeed, these local challenges would need to be addressed alongside broader international goals.

From a policy perspective, the introduction of CBDCs may also reshape how central banks manage economies. The IMF (2023) reports that CBDCs could strengthen monetary policy transmission by allowing interest rate changes to directly affect consumer wallets. They may also give governments more tools to manage capital flows and prevent currency volatility in times of crisis. These features could make CBDCs especially attractive to emerging economies looking to reduce dependence on the U.S dollar.

Overall, the development of CBDCs reflects a mix of political, technical, and economic goals. Whether for financial control or independence, CBDCs are shaping the future of money and BRICS nations may be well positioned to use them as tools for regional and global influence.

Section 3: Context

To understand whether a BRICS led digital currency could reduce dollar dominance, it's important to start with the broader picture. Central Bank Digital Currencies (CBDCs) have become one of the most active areas of interest in global finance. They're a way for governments to modernize money, increase control over transactions, and in some cases, reduce dependence on global currencies, especially the US dollar.

A CBDC is a digital version of a country's official currency, issued and managed by its central bank. Unlike decentralized cryptocurrencies like Bitcoin, CBDCs are tightly controlled and directly connected to the country's existing monetary system. They're designed to match the value of physical money and can be used for everyday payments or large scale financial transactions.

The BRICS countries Brazil, Russia, India, China, and South Africa are each exploring CBDCs in different ways. As seen in Table 1, China is furthest ahead, with its Digital Yuan rolled out in pilot programs across several cities and integrated into popular payment apps in 2020. Russia has begun public testing of its Digital Ruble, while India is experimenting with both wholesale and retail versions of the Digital Rupee which is the only BRICS nation to do so, as seen in Table 1. Brazil is focused on a wholesale system called Drex, and South Africa is involved in cross-border trials with other African countries and the Bank for International Settlements.

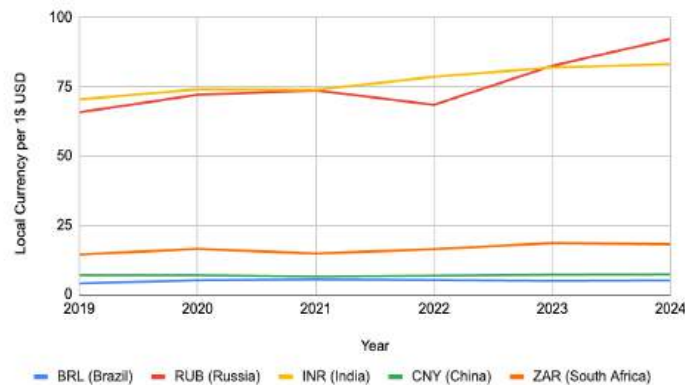
Table 1: BRICS' digital currency status

Country	Name	Year of Rolling Out	Wholesale/Retail	Exchange Rate Regime with the USD
Brazil	Drex	2023	Wholesale	Floating
Russia	Digital Ruble	2023 (in testing phrase only)	Retail	Managed float
India	Digital Rupee	2022	Wholesale & Retail	Managed float
China	Digital Yuan	2020	Retail	Managed float
South Africa	Digital Rand	2018 (in testing phrase only)	Wholesale	Managed float

Most of these CBDCs are designed to coexist with physical money and operate on a 1:1 exchange rate. So one digital rupee, ruble, or real is equal in value to its physical counterpart. This thus means that these digital currencies would have an exchange rate regime with the USD same as the physical version, which are listed in Table 1. Some systems are exploring more advanced features, like limits on transfers or programmable rules, but most are still at an early stage.

CBDCs matter in this discussion because of their potential impact on international trade and currency dependence. If BRICS countries can use CBDCs to settle transactions with each other without converting to dollars, it could reduce the global demand for dollars in cross-border trade. BRICS members don't share the same exchange rate regime, China keeps a tight grip, India manages moderately, while Brazil and South Africa float and Russia is distorted by sanctions. South Africa is the most volatile moving from 14.45 to 18.60 in just a few years whereas China remains relatively stable as the yuan is not free floating as seen in Table 1.

Graph 1: Exchange rates of BRICS Currencies vs USD (2019 - 2024)



Source: World Bank Group Official Exchange Rate

This is where CBDCs come in. The “Geopolitics of CBDCs” paper points out that strong

trade networks are a key condition for adopting shared currency infrastructure. When countries trade frequently, having a system that allows faster, cheaper, and more direct payment makes economic sense. That's part of why the euro worked in Europe. It followed years of growing trade and regulatory alignment.

BRICS isn't at that stage yet. The group has wide differences in monetary policy, currency stability, and capital controls. China's currency is tightly managed, Russia faces sanctions, South Africa's rand floats, and India controls its exchange rate to an extent. A shared digital currency would require cooperation on policy, not just technology.

This challenge is one reason the idea of a BRICS common currency has been raised. Several leaders have floated the possibility in recent years, often framed as a way to reduce reliance on the dollar in trade. The proposal is still more political than a concrete plan, but it shows the direction of thought. For now, BRICS countries are exploring digital currencies individually, and there's no sign of a fully unified system. Still, discussions about a common unit of account possibly tied to a basket of member currencies or commodities show that the bloc is at least considering deeper cooperation in the long run.

Even if BRICS countries develop strong CBDC systems, it won't immediately push out the dollar. The dollar is more than just convenient. It's backed by deep capital markets and widely trusted by institutions around the world. It serves as a store of value in uncertain times, not just a transaction tool. Most countries, including BRICS members, hold it in reserves for that reason.

Still, the direction of movement is worth paying attention to. If BRICS countries continue deepening trade ties and gradually shift to settling those trades in their own digital currencies or eventually in some shared BRICS currency it could chip away at the dollar's role. Not because of a sudden political shift, but because of practical changes in how trade works. CBDCs give BRICS countries the tools to experiment with that shift. They won't replace the dollar on their own, but they may change how value moves between nations. The trade data shows that BRICS isn't there yet, but it's getting closer. If intra-BRICS trade keeps growing and CBDC systems become more interconnected, the global financial system may start to look very different within the next decade.

Section 4: Analysis

In this section, we will analyse how a global digital currency, in particular, a BRICS digital currency, could reduce dollar dominance and shift macroeconomic power via 3 angles: 1) international trade, 2) financial services, and 3) remittance.

Section 4.1: International trade

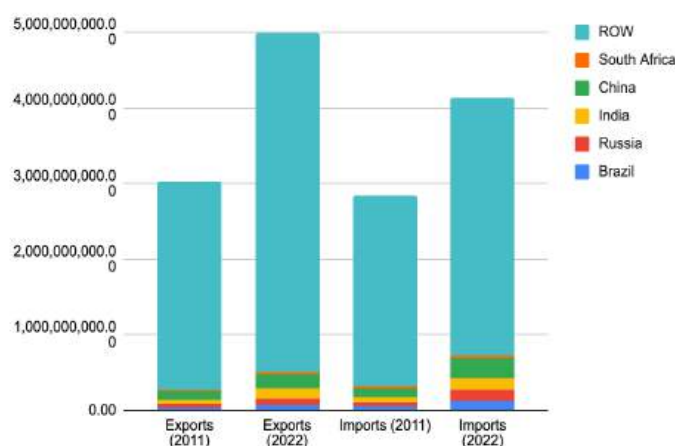
It is clear how the current international trade patterns between BRICS would call for a common BRICS digital currency, and how its existence would be able to reduce dollar dominance.

Based on trade figures from 2011 and 2022, the BRICS countries do trade with each other as seen in Graph 2, but most of their trade still goes outside the group. For example, in 2022, Brazil exported around \$334 billion worth of goods. Of that, only about \$91 billion went to other BRICS countries. The remaining \$234 billion went to the rest of the world. Its imports tell a similar story: \$292 billion total, with about \$87 billion from BRICS.

Russia's 2021 exports totaled \$492 billion, but only about \$95 billion went to BRICS. In 2011, the number was even lower, just around \$38 billion in BRICS trade. India, South Africa, and China show similar trends. In 2022, China exported over \$3.5 trillion worth of goods, yet the vast majority went outside BRICS. While intra BRICS trade has increased over time, it still makes up a minority of total trade volumes for each country.

That said, the growth in internal trade is important. From 2011 to 2022, Brazil's exports to China more than doubled. India's exports to Brazil also rose significantly. Russia's trade with China and India has grown sharply, especially following Western sanctions that pushed Russia to look elsewhere for partners. These patterns show that the BRICS trade network is strengthening, even if it isn't dominant yet.

Graph 2: Total intra- and extra- trade of BRICS nations (2011 vs 2022)



Source: World Bank WITS and author's calculation

The question then becomes how pricing trade in a shared BRICS currency would work, and what impact it could have on reducing reliance on the U.S dollar. Currently, most BRICS trade is still invoiced in dollars, meaning exporters and importers alike are exposed to exchange rate volatility and forced to pay transaction costs to convert their local currencies into dollars and then back into another national currency. For example, a Brazilian soybean exporter selling to China must typically price in dollars, exposing the transaction to dollar fluctuations even though neither Brazil nor China uses the dollar domestically. This creates inefficiency, and in times when the dollar strengthens, it erodes the competitiveness of BRICS exports on the global stage. Studies on dominant currency pricing, such as those by Gopinath et al. (2020), show how the dollar's role can amplify vulnerability by transmitting U.S monetary policy globally: when the

U.S raises interest rates, the dollar strengthens, imports for dollar using economies become more expensive, and trade contracts for exporters. A BRICS digital currency would bypass this exposure by enabling settlements in a unit they collectively govern, insulating them from at least part of the volatility associated with U.S policy.

BRICS countries face a limitation in bargaining power. Because intra-BRICS trade is a relatively small portion of their overall exports at about 10–11%, many partners outside the bloc may resist switching away from the dollar, which offers deep liquidity and broad acceptance. This data is corroborated by World Bank and IMF figures (see Graph 2), which highlight the continued reliance of BRICS nations on global dollar-based commerce.

In the short term, BRICS may not have the leverage to demand pricing in their own currency for trade with the wider world; external partners have little incentive to accept alternate units without a substantial shift in global trade patterns.

Yet there is a path forward: by consistently using a BRICS digital currency for trade within the bloc, they can normalize its use, build liquidity over time, and strengthen the internal market. Establishing a pattern of transactions in a new currency is a proven way to gradually increase trust and utility, setting the foundation for broader eventual adoption.

Already, the growth in Russia's trade with China and India after Western sanctions has demonstrated that necessity can push countries to adopt alternatives when access to the dollar is constrained. This is because maintaining stable trade flows is essential for economic resilience, so when dollar-based mechanisms are blocked, countries must innovate with new financial channels and agreements to preserve commercial continuity.

If intra-BRICS trade continues to rise, collective bargaining power will increase, making it easier to negotiate invoicing in their own unit with external partners as well. As their internal market becomes more robust, BRICS members will be better positioned to influence global trade and currency practices.

In this context, BRICS' advantage lies in its shared motivation rather than uniformity. All five members face, in different forms, vulnerability to dollar-driven shocks, whether through sanctions, exchange rate volatility, or dependency in global capital markets. This shared incentive has produced political will to cooperate where formal institutions remain limited. Unlike the Eurozone, where monetary union preceded digital innovation, BRICS' experimentation with CBDCs grew out of necessity, making it both flexible and pragmatic. This capacity to adapt policy around digital infrastructure, rather than the reverse, may be what allows BRICS to move faster in implementing alternative settlement systems.

Section 4.2 Financial Services

BRICS countries face a limitation in bargaining power. Because intra-BRICS trade is a relatively small portion of their overall exports at about 10–11%, many partners outside the bloc may resist switching away from the dollar, which offers deep liquidity and broad acceptance. This data is corroborated by World Bank and IMF figures (see Graph 2), which highlight the continued reliance of BRICS nations on global dollar based commerce.

In the short term, BRICS may not have the leverage to demand pricing in their own currency for trade with the wider world; external partners have little incentive to accept alternate units without a substantial shift in global trade patterns.

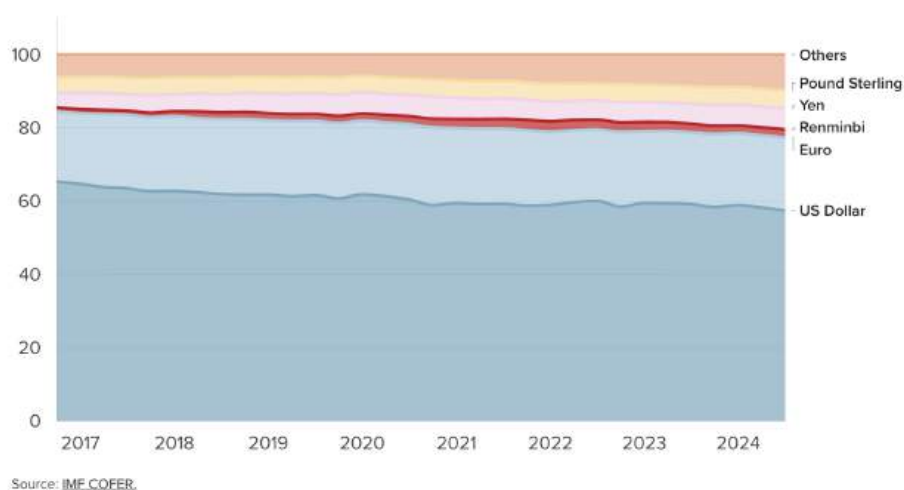
Yet there is a path forward: by consistently using a BRICS digital currency for trade within the bloc, they can normalize its use, build liquidity over time, and strengthen the internal market. Establishing a pattern of transactions in a new currency is a proven way to gradually increase trust and utility, setting the foundation for broader eventual adoption.

Already, the growth in Russia's trade with China and India after Western sanctions has demonstrated that necessity can push countries to adopt alternatives when access to the dollar is constrained. This is because maintaining stable trade flows is essential for economic resilience, so when dollar based mechanisms are blocked, countries must innovate with new financial channels and agreements to preserve commercial continuity.

If intra-BRICS trade continues to rise, collective bargaining power will increase, making it easier to negotiate invoicing in their own unit with external partners as well. As their internal market becomes more robust, BRICS members will be better positioned to influence global trade and currency practices.

A BRICS digital currency has transformative potential for the cross border financial services landscape within the bloc. Currently, even when financial flows such as loans, portfolio investments, interbank transfers, and corporate transactions occur between BRICS entities, the U.S dollar remains the principal currency of settlement as it accounts for roughly 58% of global official foreign exchange . As seen in Graph 3, the U.S dollar still makes up the majority of reserves, though its share has been gradually decreasing, while the renminbi remains small but is slowly rising. This reliance persists despite both counterparties using local BRICS currencies domestically, and is sustained by the deep liquidity and historical dominance of the dollar in global finance. If BRICS nations begin to price and settle their intra bloc financial services in a common digital currency, several important shifts would follow.

Graph 3: Currency composition of official foreign exchange reserves (%)



The exposure of BRICS institutions to dollar volatility would diminish, removing the inefficiencies and costs tied to swapping assets into and out of dollars for each transaction. This reduction in foreign exchange risk and conversion costs would be particularly impactful, as financial services contracts are often structured on tight margins where repeated small cost savings accumulate significantly over time. Moreover, the operational efficiency of these markets would rise; financial services frequently require instant, high value transactions where the ability to process and settle in real time something a well designed digital currency can provide would create a marked advantage. Current intra BRICS financial flows are growing but still constitute a minority of each country's overall external finance. Chinese banks' increasing activity in Russia, Brazil, and South Africa, along with Indian conglomerates expanding their cross border financial dealings within the bloc, signal that the framework for deeper intra BRICS financial integration is being built, but the underlying flows remain dwarfed by transactions with partners outside the group.

Similar to international trade, a key question emerges around the bloc's bargaining power to insist that external parties also settle in a BRICS currency when dealing with financial transactions. Right now, because intra group flows are still relatively small, the leverage to demand this is limited. By comparison, the US dollar currently represents 58% of global foreign reserve holdings, while the euro accounts for only 20 percent. Together, the dollar and euro make up nearly 80 percent of global foreign exchange reserves, revealing the scale of dominance that BRICS would be seeking to challenge.

However, if the group can foster critical mass in internal usage, liquidity and trust in the currency will deepen, setting the stage for gradual internationalization. By requiring or strongly encouraging BRICS-based financial services to settle in the new digital unit, the group could normalize its use, build technical infrastructure, and develop a track record of stability and transparency. Over time, these steps would make it progressively easier to convince, or eventually require, external institutions to transact in the BRICS currency when dealing with the bloc.

Still, it is important to consider whether a BRICS CBDC could simply replicate some of the same issues the dollar system created. While a digital framework would improve transparency and reduce conversion costs, it could also concentrate power in central authorities if not properly designed. Countries like China or Russia, with more centralized financial systems, might dominate transaction flows, raising concerns about governance balance. However, a shared digital currency structured around distributed infrastructure and multilateral oversight could help prevent that outcome, ensuring that efficiency does not come at the cost of neutrality.

The benefits of implementing a CBDC in this context would be substantial. Such a system allows immediate settlement, lowers costs, strips away much of the friction tied to legacy cross currency systems, and offers BRICS greater resilience from dollar-centric monetary shocks. Because the notional value of international financial flows is typically much larger per transaction than that of goods trade, even limited migration from dollar to BRICS digital currency settlement could have an outsized impact on regional and global financial equilibrium.

Section 4.3 Remittance

The sphere of remittances provides another powerful argument for the adoption of a BRICS digital currency. Large populations from India, China, South Africa, Brazil, and Russia live and work in each other's territories and around the world, generating significant cross border flows each year. According to the World Bank, remittance flows to low and middle income countries (LMICs) which prominently include BRICS members, are projected to reach a record \$690 billion by the end of 2025. Of that sum, BRICS countries play a central role: in 2025, India alone is expected to receive over \$120 billion, with China and Brazil also ranking among the top global recipients.

The scale of remittance among BRICS nations is massive and persistent, with India alone receiving close to \$135 billion per The World Bank and Reserve Bank India (RBI). Though the exact numbers for intra-BRICS remittances are less readily available, there is no doubt about the depth and regularity of these flows. Historically, remittance companies needed to maintain liquidity and currency reserves for every major BRICS currency, creating unnecessary overhead and complications. A single regional digital currency would allow these companies to operate more efficiently and cost effectively, making the case to prioritize the BRICS unit over the dollar for many of their internal and regional transactions. The combined bargaining power of a bloc unified in its use of a common digital currency would enable remittance companies and their customers alike to realize substantial savings and process flows more quickly, bypassing the old currency paradigm and challenging the incumbent dominance of the dollar.

A CBDC adds value in this context by supporting instant, trackable, and programmable transfers between individuals and firms, ensuring compliance while reducing settlement time to near zero. For remittances, which tend to be frequent and of relatively modest size, smoothing out the cumulative costs of transaction and conversion is critical. The adoption of a BRICS CBDC would offer not just lower fees, but also greater transparency, security, and ease of regulatory oversight. For the firms operating these networks, the technology investment required would be lower if focused on a single standard rather than five. In the aggregate, this shift would promote financial inclusion, empower transnational families, and further erode the transactional supremacy of the U.S dollar in a sector of economic life that is both vast and deeply human.

Section 5: Conclusion

This paper examined the potential of a BRICS central bank digital currency (CBDC) to reduce dollar dominance through three main channels: international trade, financial services, and remittances. These angles were selected because they represent the largest and most persistent cross border flows in the global economy. Trade volumes reveal how much of BRICS' activity still relies on the dollar, despite growth in intra bloc exchanges. Financial services highlight the depth of dollar entrenchment, where cross border loans, investments, and interbank transfers remain overwhelmingly dollar denominated. Finally, remittances demonstrate the human scale of dollar dependence, where millions of families across BRICS absorb the costs of conversion and fees. Taken together, these perspectives show how a shared BRICS CBDC could gradually chip

away at the structural supremacy of the dollar.

Nevertheless, it is essential to acknowledge that CBDCs are not a perfect substitute for traditional systems. They introduce new questions around privacy, interoperability, and governance. If dominated by state-controlled frameworks, a BRICS CBDC could risk concentrating monetary influence rather than democratizing it. Yet if designed collaboratively, with transparent settlement protocols and mutual oversight, it could create a fairer and more balanced financial network.

Ultimately, BRICS may be particularly well-suited for this transition because of its shared strategic incentives and the complementary nature of its economies. Each member gains from reducing dollar exposure, and all face similar pressures to strengthen monetary sovereignty. A cooperative CBDC model allows them to pursue this goal collectively while retaining national control. In this way, BRICS offers a realistic testing ground for digital monetary coordination, less integrated than the EU, but more aligned in intent than most regional blocs.

The paper also considered the broader context of BRICS' position in the global economy, where the dollar accounts for about 58% of official foreign exchange reserves compared with less than 3% for the renminbi. Among the three channels analyzed, trade appears to be the most powerful in the short term because of its scale and frequency, and because exporters and importers are highly sensitive to transaction costs and currency volatility. Financial services, by contrast, are slower to shift because of their complexity and integration with global markets, but they carry significant weight given the sheer value of financial flows. Remittances, while fragmented across millions of smaller transactions, are socially vital and politically relevant, especially for India, which receives over \$120 billion annually. In this way, trade demonstrates the clearest pathway to scale, financial services the deepest structural challenge, and remittances the strongest argument for inclusion and accessibility.

Going forward, policymakers should prioritize building technical and institutional frameworks that make trade settlement in a BRICS CBDC seamless. Governments and central banks need to ensure connectivity between their domestic payment systems, while also creating incentives for firms, particularly small and medium sized exporters to adopt the new currency. For external partners that already trade heavily with BRICS but are not tied strongly to the dollar, such as commodity exporters in Africa and Latin America, bilateral agreements could help expand the reach of the CBDC. Financial regulators should also explore ways to denominate cross border financial contracts in the new unit, while remittance corridors could be strengthened with digital solutions that reduce costs for households. A coordinated approach across these fronts would maximize both adoption and credibility of the new system.

This paper therefore reframes the question not as whether BRICS will replace the dollar, but whether digital integration within the bloc can make global finance more plural and less vulnerable to the dominance of any single currency. The answer lies not only in technology, but in cooperation.

There are, however, limitations to this paper. By focusing on trade, financial services, and remittances, it excluded other channels such as foreign direct investment, tourism, and debt

markets, which may also shape the trajectory of a BRICS CBDC. In addition, the three angles were analyzed in isolation, even though in reality they are interconnected trade flows that influence financial systems, and remittance patterns depend on both. Treating them separately may underestimate the true multiplier effects of CBDC adoption. Despite these constraints, the paper demonstrates that a BRICS CBDC offers more than just a technical shift; it is a political and economic tool that could rebalance global finance. Ultimately, while the dollar is unlikely to be displaced overnight, the development of a BRICS CBDC provides an important first step toward diversifying the monetary order and giving emerging economies greater autonomy in shaping the rules of global commerce.

Works Cited

- Bank for International Settlements (BIS), et al. (2021). Central bank digital currencies: Financial stability implications. BIS Publications.
- Bindseil, U., Panetta, F., & Terol, I. (2020). Central bank digital currency: Functional scope, pricing, and control. European Central Bank Occasional Paper Series, No. 247.
- Coquidé, C., Lages, J., & Shepelyansky, D. L. (2023). Emergence of a new world monetary order. *Physica A: Statistical Mechanics and its Applications*, 616, 128673.
- Economic Times. (2025, June 30). Diaspora remittances hit a new record at \$135.46 bn in FY25. The Economic Times.
<https://economictimes.indiatimes.com/nri/invest/diaspora-remittances-hit-new-record-at-135-46-bn-in-fy25/articleshow/122144367.cms>
- Eren, E., & Malamud, S. (2019). Dominant currency debt. BIS Quarterly Review, December.
- Federal Reserve. (2025, February 27). Global remittances cycle. Board of Governors of the Federal Reserve System.
<https://www.federalreserve.gov/econres/notes/feds-notes/global-remittances-cycle-20250227.html>
- FinTech Biz News. (2024, June 27). Remittances may grow to \$690 bn by 2025, India \$120 bn top recipient for remittances. FinTech Biz News.
<https://www.fintechbiznews.com/fintech-digital-payment-/remittances-likely-to-be-up-to-690-bn-by-2025-india-120-bn-top-recipient-for-remittances>
- Gopinath, G., Boz, E., Casas, C., Díez, F. J., Gourinchas, P. O., & Plagborg-Møller, M. (2020). Dominant currency paradigm. *American Economic Review*, 110(3), 677–719.
- International Monetary Fund (IMF). (2023). The rise of central bank digital currencies. IMF Policy Paper.
- International Monetary Fund (IMF). (2025, July 17). Currency composition of global official foreign exchange reserves. IMF Data.
https://data.imf.org/en/news/imf%20data%20brief%20jul%209?utm_
- Nottingham Trent University. (2024). The geopolitics of CBDCs. Centre for Policy and International Relations.
- Nugraha, A. (2024). Barriers to CBDC adoption in developing economies. *Journal of Emerging Financial Technology*, 12(1), 55–78.
- Pettis, M. (2023). Trade wars are class wars: The economic impact of dollar dominance. Princeton University Press.
- Subacchi, P., et al. (2021). The geopolitics of digital currencies. Chatham House Research Paper.
- World Bank. (2023). Migration and development brief 38: Remittances brave global headwinds. Washington, DC: World Bank Group.
- World Bank Group. (2024). Official exchange rate data. World Bank Data.
<https://www.worldbank.org>
- Atlantic Council. (2025, January). Dollar dominance monitor. Atlantic Council GeoEconomics Center.
<https://www.atlanticcouncil.org/programs/geoeconomics-center/dollar-dominance-monitor>

Exploring the Effects of Stress, Anxiety, and Depression on Cognition, and the Therapeutic Influence of Music By Tamira Darisala and By Mentored by Dr. Deblina Dey

Abstract

Stress, anxiety, and depression are highly interconnected psychological conditions that can affect cognitive processes, emotional regulation, and even overall health. Stress is the body's physiological response to external environmental demands exceeding the organism's capacity to cope with the external environment, it influences brain structures such as the amygdala and the prefrontal cortex, altering memory and emotional processing. Anxiety disorders are among the most prevalent psychological issues globally and depression which is characterized by persistent negativity and low motivation usually occur together closely and both impair concentration, working memory and attention. Chronic stress often exacerbates these issues by elevating stress hormones such as glucocorticoids which can cross the blood brain barrier and disrupt neural function. This review highlights how music can serve as a therapeutic method, and how it positively affects brain chemistry and physiology to reduce the symptoms of these mental health issues. Evidence found in both human and animal studies shows that music stimulates the release of dopamine, serotonin and oxytocin, decreases cortisol levels and can enhance positive neural activity in regions of the brain responsible for critical memory and emotion processing. Clinical studies confirm this treatment's efficiency as well as the level of accessibility. By addressing both the physiological and psychological aspects of mental health, music can emerge as a powerful, noninvasive and accessible treatment for supporting cognitive recovery in individuals affected by these mental health conditions.

Introduction

Stress, anxiety, and depression are three interrelated psychological conditions that significantly influence individual functioning or mental health outcomes. Stress scientifically is when the body responds to stimulus in a way where it can change bodily functions such as blood pressure, heart rate, and blood sugar levels. However the term stress has been used excessively to the point at which its range is incredibly vast and this creates problems when narrowing down studies. A proposed solution is to restrict the term to when an external environmental condition exceeds the natural capacity of an organism (Koolhaas, et al.). This is because scientific stress can be categorized in two ways: an unpredictable response which is anticipatory or an uncontrollable response that is a reduced recovery (Koolhaas, et al.).

Anxiety disorders are the most common mental health issue in the western world and significantly impact people's ability to function. (Wittchen et al.) severely impairing individuals ability to function as well as the basic quality of life. Anxiety is characterized by persistent and excessive worry, fear, or nervousness, that is often in response to facing uncertainty. These disorders generalized anxiety, social anxiety, separation anxiety, and specific phobia disorders affect approximately 374 million people worldwide or 4802 cases per 100,000 in the population (Liu et al.).

Depression is a mood disorder marked by prolonged sadness, a loss of interest and reduced motivation. Eligible symptoms for depressive episodes include the additional symptom of the loss of confidence or self esteem. (Paykel.). While it is different from anxiety in terms of diagnostics, the two share similar diagnosing characteristics such as sleep disturbances, concentration difficulties, and negative thought patterns. Prolonged exposure to stress can be a major contributing factor to both anxiety and depression, further highlighting the interconnectedness of these conditions.

Current Treatments and Side Effects of Antidepressants

Depression and Anxiety are often treated with medication, but the types of drugs and treatments used can come with benefits as well as potentially serious side effects. Doctors weigh the potential improvements in the condition against the risks that each drug carries. Three common classes of medications include SSRIs/SNRIs, benzodiazepines and tricyclic antidepressants(TCAs). These medications work differently in the brain but can all still affect people in both positive and negative ways (Mayo Clinic).

Selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors(SNRIs) are usually considered initial treatments for depression because they are safer than older drugs. They work by balancing serotonin and norepinephrine, however these drugs can cause problems and side effects. Emotional blunting, where patients may feel less sadness but also less joy, leaving the patient emotionally numb. Withdrawal symptoms can also happen if someone stops the medication abruptly, including dizziness, headaches, and mood swings. A greater challenge is that scientists still have limited long-term safety data, as there isn't enough scientific evidence to suggest what will happen with decades of use of these treatments to the body and brain.

Benzodiazepines are prescribed for anxiety and agitation that can come with depression though they are not the first choice typically. This treatment calms the nervous system but the risks are high. Patients can develop dependence and addiction quickly, making it difficult to discontinue without medical intervention. Older adults who take Benzodiazepines can face a higher chance of accidents and falls due to drowsiness and slower reflexes caused by drugs. Some patients even experience paradoxical reactions meaning instead of calming them the drug can increase agitation and even aggression. There is evidence to also suggest there may be a higher mortality risk with long-term benzodiazepine use (Mayo Clinic).

Tricyclic antidepressants are one of the oldest classes of antidepressant medications. They are effective however less commonly prescribed today due to their heavy side effect profile. They often cause sedation, which can interfere with daily functioning, and weight gain, which can lead to other health issues such as diabetes and heart disease as well as cardiac risk since TCAs can affect heart rhythm and blood pressure which makes them dangerous for people with heart conditions. They also have high overdose potential, making even the smallest amount above the prescribed dose life threatening making them especially risky for people struggling with suicidal thoughts.(Mayo Clinic).

While antidepressant medications are incredibly important for the treatment and wellbeing of many people, they can carry significant side effects that must be carefully managed. This is why treatment decisions should be heavily individualized and taken with great medical care, while other non drug treatments should be studied as a supplement or complement to these antidepressants.

Effect of Stress on memory and the Amygdala

The amygdala is responsible for processing emotions and containing emotional memories, therefore stress can have a deep influence on the memories stored and how the amygdala functions. Emotionally stimulating events and memories (positive and negative) are remembered with more detail than neutral ones that don't contain any significant stimulus, and the amygdala plays a central role in this process (Hamann et al.). Studies with humans and experimental animals have demonstrated that memory is better for emotionally arousing stimuli than for emotionally neutral stimuli (Hamann et al.). This emotional clarification of the memory is closely tied to stress hormones such as cortisol and epinephrine (Mayo Clinic). In animal studies it has been found that the amygdala is essential in strengthening long term memory of emotional events by working with peripheral adrenergic systems within the amygdala. (Hamman et al.). When the body experiences stress, it releases hormones that stimulate and activate the amygdala, which then strengthens the memory and clarity of the stressful event (Hamman et al.). Roozendaal and McGaugh 2009 also emphasize this as the amygdala's ability to adapt in both structure and function makes it key to understanding how emotional and fearful experiences are remembered, as well as how long term and chronic stress can lead to harmful effects. The basolateral amygdala specifically is sensitive to stress hormones and works together with other brain regions like the hippocampus to store emotional memories (Kim et al.). This shows how stress doesn't just cause short emotional reactions, it also changes how the brain strengthens and remembers experiences. Short term stress triggers the release of various hormones and neurotransmitters, many of which play a role in shaping how memories are stored, with evidence that the basolateral amygdala is a central area where memory consolidation comes together (Roozendaal et al.).

Cognition, defined as the set of mental processes involved in knowledge and understanding through thought, experience, and the senses, plays a critical role in the context of these disorders.(Kirkanski et al.). The two essential parts of cognition are learning- the process of encoding new information and memory which is the storage and retrieval of information. Both anxiety and depression can affect cognition with the symptoms possibly being long lasting. High levels of anxiety can be associated with reduced working mental capacity and attention span while depression can affect processing speed and memory retrieval. Stress can affect cognition by a great amount. Both external (exogenous) and internal (endogenous) increases in stress hormones can alter cognitive performance. These stress hormones could thus cross the blood brain barrier and access the brain where they could influence cognition and mental health(Lupien et al.). Hormones such as glucocorticoids can cross the blood brain barrier and reach areas such

as the frontal cortex and the amygdala. These parts of the brain are important as they function with helping in memory and emotion. Originally used as anti-inflammatory medications, glucocorticoids were found to cause ‘steroid psychosis’, revealing that they can directly affect brain function and mental health (Lupien et al.). Stressful situations cause the body to release these hormones which can eventually impair learning and memory. As stress is closely linked to both anxiety and depression, prolonged exposure to stress can worsen emotional health and contribute to the development of these disorders, further amplifying its negative effects on memory and the brain.

Effect of anxiety and depression on memory

Depression has a powerful impact on how people think, focus, and remember. People with depression often struggle with concentration and memory, while at the same time becoming stuck in cycles of negative thinking as depressed people often report having difficulty involving concentration and memory (Matthews & Macleod). With these struggles, depressed individuals tend to stay stuck on the negative aspects of life, these depressed individuals report easily concentrating on the negative self thoughts and they exhibit enhanced recall negatively enhanced memories, information, and events while experiencing this depressive state of mind (Matthews & MacLeod.). This combination of lack of concentration and negativity can make it very difficult for individuals with depression to break out of dangerous and harmful thought patterns.

One explanation for these cognitive difficulties is the resource allocation hypothesis, which suggests that depressed individuals, because their cognitive capacity is reduced therefore deficits can be found in remembering and engaging in other cognitive processes (Joormann). Similarly, the affective interference hypothesis states that because depressed individuals have deficits with processing of emotional material, their performance will suffer and falter with tasks that require emotional ignorance on aspects of the stimuli and respond to other aspects of the material (Joormann). Therefore depression consumes mental resources and draws focus toward emotional information, especially negative content. These theories are supported by findings that depressed individuals often show memory problems when tasks require more focus, Impairments stemming from depression are not observed in all components of memory but they are primarily found in tasks including recall and recognition (Joormann). In less structured situations, depressed individuals are more likely to contemplate and second guess, which interferes with memory and problem solving.

Depression is also linked to attention and memory biases, while depressed people do not always show immediate biases toward negative behavior, once this information has their attention, they have difficulties disengaging and disregarding it (Joormann). This difficulty letting go of negative thoughts and images often leads to second guessing and extra contemplation, Depression is associated with the expansion of and difficulties disengaging from a negative environment (Joormann). This eventually makes it harder for depressed people to focus on new information or regulate their emotions.

Overall, depression does not always cause cognitive impairments, but it does make focusing and remembering harder in some situations, specifically when negative thoughts are involved. Depression strongly affects cognition by reducing concentration, impairing memory in certain contexts, and creating powerful biases toward negative thoughts. These effects often trap individuals in cycles of second guessing, which not only maintains depression but can also make recovery more difficult.

Effect of music on Stress and Depression

Music affects stress in many ways due to how it changes the brain's chemistry and structure. When music is listened to or played, the brain releases chemicals like dopamine and serotonin. These are neurotransmitters that are linked to feeling good and calm. Music is able to change the chemistry in the brain, inducing the release of neurotransmitters, namely dopamine and serotonin, that activate the reward system and hormone release (Salimpoor et al.). This means that just listening to music can activate the same brain systems that make us feel happy or rewarded. Music also affects oxytocin, the hormone that helps us bond with other people and feel emotionally connected. Musical therapies in groups and singing can increase the production of oxytocin found in the brain (Chanda and Levitin). This is the reason why singing with others or attending musical events such as concerts can make people feel less stressed and more connected. (Hansen and Keller) also pointed out that oxytocin appears key to establishing and maintaining social bonds therefore allowing for music's social bonding effects. Therefore music isn't just something we hear, it also helps us feel closer to others by creating bonds and connectedness, which can reduce stress. Music also creates physical changes in our bodies that are directly tied to stress, music induces a wide range of physiological effects on the human body, including changes in respiration, heart rate, skin conductivity, blood pressure, skin temperature, muscle tension, and biochemical responses (Zaatar et al.) These are all reactions that the body has when it relaxes, which shows that music can physically calm the biological systems in the body down. The areas of the brain that process reward and emotion, such as the amygdala, are also activated when we hear music. The emotional response to music was associated with increased blood flow in the brain areas associated with reward and emotion (Blood & Zatorre). This means music actually reaches the emotional centers of our brain that are affected by stress. These same brain areas are used when people experience pleasure from other things like food, social contact, or even drugs, according to (Speranza et al.).

Music therapy, in particular, has been shown to significantly improve anxiety and stress-related symptoms in a variety of populations (Guetin et al., Horne-Thompson and Grocke). Because of its noninvasive features and easy accessibility, music therapy can now be an alternative or supplementary treatment for psychiatric conditions (Lin et al.). As an example, (Goldbeck and Ellerkamp) found that music therapy helped children with anxiety reduce symptoms and improve their quality of life, while Lin et al. also showed strong evidence that it reduces anxiety in cancer patients.

Music can make a major difference in reducing depression and anxiety, both through

improving human behavior and changing brain chemistry. In studies that focus on animals, the effects of music were especially clear when researchers looked at how it helped mice exposed to chronic unpredictable mild stress (CUMS). These stressed mice usually showed signs of depression, such as not moving much or avoiding open spaces during experiments. However, the mice that listened to music during the experiment were far more active and engaged. They did not exhibit despair, whereas CUMS mice exhibited severe behavioral despair and overall did not exhibit depression-like or anxious behavior, even when subjected to high levels of stress (Fu, Qiu, Chen, Chen, Qi, & Cheng.). This shows that even under high levels of stress, music can reduce depression-like behaviors.

Music also had an effect on the survival of neurons in the hippocampus, a part of the brain important for regulating mood and memory. Stress caused a large decrease in the number of neurons in this region, while the music group showed no significant reduction compared to normal mice (Fu et al.). The researchers also looked at proteins connected to whether cells survive or die. They found that stressed mice had lower levels of a protective protein called Bcl-2 and higher levels of Bax, which promotes cell death. However, in the music group, the balance of these proteins was close to that of the healthy brain, which means that music not only helps with symptoms but also protects the actual structure of the brain from damage linked to depression.

These positive effects are not limited to animal studies. Human research shows that music can also make a big difference for patients facing serious challenges, like breast cancer. One meta-analysis looking at 13 different studies found that music therapy can significantly improve anxiety in breast cancer patients while also reducing depression (Xu, Liu, Fan, Li, & Li). The study also found that certain types of interventions worked best. Patients experienced the strongest results when they listened to their favorite music for less than 30 minutes, twice a day, for a cycle of two to three weeks. In fact, shorter sessions were more effective than longer ones, possibly as people could focus better and stay engaged when the music time was not too long.

Researchers also explained why music has these effects on humans. Listening to music can change hormone levels in ways that benefit mental health, as it can increase the levels of estrogen and oxytocin, and also reduce the concentration of serum cortisol in women, so that the level of cortisol in breast cancer patients can return to normal (Xu et al.). Since high cortisol is connected to high stress and depression, lowering it helps patients feel calmer. Music also impacts the brain's reward system by boosting dopamine activity in areas like the nucleus accumbens and ventral tegmental cortex, which are linked to motivation and pleasure. In addition, it influences brain structures such as the hypothalamus and insula, helping regulate emotional responses. These changes allow patients to better control anxiety and depression.

Altogether, music can improve depression and anxiety by working on both the mind and the body. Animal studies prove that it reduces despair and depression-like behaviors, lowers harmful inflammation, and even protects neurons in the brain, while human studies confirm that it helps patients with serious illnesses manage their symptoms more effectively. With its ability

to heal both emotionally and biologically, music is a very effective and powerful form of therapy for those struggling with depressive behavior.

Summary

Stress, anxiety, and depression are interconnected psychological conditions that can significantly impact cognition, emotional regulation, and overall health. Stress is the body's physiological response to environmental demands exceeding an organism's capacity, affecting brain structures such as the amygdala and prefrontal cortex. This can alter memory, attention, and emotional processing (Koolhaas et al.). Anxiety disorders, characterized by excessive worry and fear, are among the most prevalent mental health issues worldwide, while depression involves prolonged sadness, low motivation, and constant negative thinking patterns (Wittchen et al., Paykel,). Chronic stress exacerbates both conditions by increasing glucocorticoid levels, which can cross the blood-brain barrier and disrupt neural function.

Stress affects memory primarily through the amygdala, which strengthens emotionally charged memories, particularly negative ones. The amygdala works with the hippocampus and other brain regions to consolidate emotional memories, and chronic stress can impair cognitive processes such as attention, working memory, and learning (Hamann et al., Roozendaal & McGaugh, 2009; Kim et al.). Anxiety and depression further impair cognition. Depression, for example, reduces concentration, biases attention toward negative information, and strengthens recall of negative memories, often trapping individuals in cycles of rumination and second-guessing (Mathews & MacLeod, Joormann,).

Music has emerged as an effective therapeutic intervention for stress, anxiety, and depression. Listening to music stimulates the release of dopamine, serotonin, and oxytocin while reducing cortisol, thereby improving mood, social bonding, and emotional regulation (Salimpoor et al., Chanda & Levitin,). Music also influences physiological systems, lowering heart rate, blood pressure, and muscle tension, while activating reward and emotion centers in the brain, including the amygdala and hippocampus (Blood & Zatorre, Zaatar et al.). Animal studies show that music reduces depression-like behaviors, preserves hippocampal neurons, and balances proteins critical for cell survival (Fu et al.). Human studies, including those with breast cancer patients, confirm music therapy's ability to reduce anxiety and depression, particularly when short, regular sessions of preferred music are used (Xu et al.).

Overall, stress, anxiety, and depression interact to impair cognition, memory, and attention, while music therapy offers a noninvasive, accessible intervention that supports emotional regulation and cognitive function. By improving both brain chemistry and physiology, music can help individuals recover from the cognitive and emotional challenges associated with these mental health conditions.

Future Directions

While current research demonstrates the overall effectiveness of music therapy in reducing stress, anxiety, and depression, and improving cognition, several areas can be explored

further. Long term human studies can be done to examine the effects of musical interventions on brain structure, emotional regulation, and memory performance. Figuring out specifics on which styles, genres, frequencies, and specific instruments influence cognition and emotional outcomes the most can optimize therapy protocols. Furthermore as depression, anxiety, and stress diagnoses skyrocket in the world it is imperative to find how musical therapy compares to other noninvasive interventions. This could potentially determine whether music has unique benefits when combined with other treatments. Additionally, this research would be best utilized focusing on the vulnerable populations in the world including children, older adults, and patients with chronic illness. By pursuing these directions, future research can enhance our understanding of how music supports mental health, cognitive recovery, and overall well being.

Works Cited

- Arnsten, Amy F. T., et al. "The Effects of Stress Exposure on the Prefrontal Cortex: Translating Basic Research into Successful Treatments for Post-Traumatic Stress Disorder." *Neurobiology of Stress*, vol. 1, no. 1, 2015, pp. 89–99. Elsevier, <https://doi.org/10.1016/j.ynstr.2014.10.002>.
- Basudan, Sarah, Nouf Binanzan, and Aesha Alhassan. "Depression, Anxiety and Stress in Dental Students." *International Journal of Medical Education*, vol. 8, 2017, pp. 179–86. IJME, <https://doi.org/10.5116/ijme.5910.b961>.
- Blood, Anne J., and Robert J. Zatorre. "Intensely Pleasurable Responses to Music Correlate with Activity in Brain Regions Implicated in Reward and Emotion." *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, 2001, pp. 11818–23. National Academy of Sciences, <https://doi.org/10.1073/pnas.191355898>.
- Chanda, Mona Lisa, and Daniel J. Levitin. "The Neurochemistry of Music." *Trends in Cognitive Sciences*, vol. 17, no. 4, 2013, pp. 179–93. Elsevier, <https://doi.org/10.1016/j.tics.2013.02.007>.
- Fu, Qiang, et al. "Music Prevents Stress-Induced Depression and Anxiety-Like Behavior in Mice." *Nature Communications*, 2023. NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC10570293/pdf/41398_2023_Article_2606.pdf.
- Goldbeck, Lutz, and Tanja Ellerkamp. "A Randomized Controlled Trial of Multimodal Music Therapy for Children with Anxiety Disorders." *Journal of Music Therapy*, vol. 49, no. 4, 2012, pp. 395–413. Oxford Academic, <https://doi.org/10.1093/jmt/49.4.395>.
- Gotlib, Ian H., and Jutta Joormann. "Cognition and Depression: Current Status and Future Directions." *Annual Review of Clinical Psychology*, vol. 6, 2010, pp. 285–312. Annual Reviews, <https://doi.org/10.1146/annurev.clinpsy.121208.131305>.
- Horne-Thompson, Annette, and Denise Grocke. "The Effect of Music Therapy on Anxiety in Patients Who Are Terminally Ill." *Journal of Palliative Medicine*, vol. 11, no. 4, 2008, pp. 582–90. Mary Ann Liebert, <https://doi.org/10.1089/jpm.2007.0193>.
- Kircanski, Katharina, Jutta Joormann, and Ian H. Gotlib. "Cognitive Aspects of Depression." *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 3, no. 3, 2012, pp. 301–13. Wiley Online Library, <https://doi.org/10.1002/wcs.1177>.
- Koolhaas, Jaap M. "Stress Revisited: A Critical Evaluation of the Stress Concept." *Neuroscience & Biobehavioral Reviews*, vol. 35, no. 5, 2011, pp. 1291–301. Elsevier, <https://doi.org/10.1016/j.neubiorev.2011.02.003>.
- Krishnamurthya, T., and B. Venkanna Rao. "Indian Classical Mohana Raga (Instrumental Music) Overcomes Anxiety, Depression and Memory Impairment in Chronic Unpredictable Mild Stress Rat Model." *Indian Journal of Traditional Knowledge*, 3 Apr. 2025. CSIR-NIScPR, <https://or.niscpr.res.in/index.php/IJTK/article/view>.
- Leubner, Daniel, and Thilo Hinterberger. "Reviewing the Effectiveness of Music Interventions in Treating Depression." *Frontiers in Psychology*, vol. 8, no. 1109, 2017. Frontiers, <https://doi.org/10.3389/fpsyg.2017.01109>.

- Liu, Chang H., et al. "Factors Associated with Depression, Anxiety, and PTSD Symptomatology during the COVID-19 Pandemic: Clinical Implications for U.S. Young Adult Mental Health." *Psychiatry Research*, vol. 290, 2020, p. 113172. Elsevier, <https://doi.org/10.1016/j.psychres.2020.113172>.
- National Cancer Institute. "Stress." *NCI Dictionary of Cancer Terms*, 2 Feb. 2011, <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/stress>.
- Salimpoor, Valorie N., et al. "Anatomically Distinct Dopamine Release during Anticipation and Experience of Peak Emotion to Music." *Nature Neuroscience*, vol. 14, no. 2, 2011, pp. 257–62. Nature, <https://doi.org/10.1038/nn.2726>.
- Song, Yao, Numan Ali, and Urs M. Nater. "The Effect of Music on Stress Recovery." *Psychoneuroendocrinology*, vol. 168, 2024, p. 107137. Elsevier, <https://doi.org/10.1016/j.psyneuen.2024.107137>.
- Speranza, Luigi, et al. "Music Affects Functional Brain Connectivity and Is Effective in the Treatment of Neurological Disorders." *Reviews in Neurosciences*, 2022. De Gruyter, <https://doi.org/10.1515/revneuro-2021-0135>.
- Wittchen, Hans-Ulrich, et al. "The Size and Burden of Mental Disorders and Other Disorders of the Brain in Europe 2010." *European Neuropsychopharmacology*, vol. 21, no. 9, 2011, pp. 655–79. Elsevier, <https://doi.org/10.1016/j.euroneuro.2011.07.018>.
- Xu, Ziqi, et al. "Preparing to Download..." *Scientific Reports*, 2025. NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC11255342/pdf/41598_2024_Article_66836.pdf
- Zaatar, Mohammad T., et al. "The Transformative Power of Music: Insights into Neuroplasticity, Health, and Disease." *Brain, Behavior, & Immunity – Health*, vol. 35, 2023. Elsevier, <https://doi.org/10.1016/j.bbih.2023.100716>.

Ultraviolet Vision in *Heliconius* Butterflies: From Opsin Duplication to Behavioral Ecology

By Max Gan

Abstract

Most butterfly lineages are not capable of true ultraviolet(UV) color discrimination. *Heliconius* butterflies are a notable exception, with some lineages possessing two distinct UV-sensitive opsins. This review synthesizes recent advances in our understanding of *Heliconius* UV vision, highlighting how an ancestral opsin duplication drives behavioral adaptations in foraging and mate selection. Altogether, members of the *Heliconius* genus are important model organisms for linking evolution to behavioral ecology.

Introduction

Animal visual systems evolve through ecological and social pressures, balancing physiological constraints with the need to survive and reproduce. Key selective pressures include foraging efficiency, predator avoidance, and mate recognition (Gerl & Morris, 2010). In insects, UV sensitivity is common, but the ability to detect differences in UV wavelengths is rare, making *Heliconius* butterflies a standout model for investigating true UV color discrimination and its ecological consequences.

This review compiles perspectives from genetics, physiology, behavior, and comparative biology to address three topics: 1. How did the unique visual system of *Heliconius* butterflies arise and how is it regulated? 2. How does UV perception influence foraging, mating, and host plant selection? 3. What do *Heliconius* butterflies reveal about the future applications and broader significance of sensory research? Through connecting mechanisms at different biological levels, this paper illustrates how genetic changes can enable novel ecological strategies, shedding light on mechanisms behind sensory evolution.

The Molecular Basis of UV Vision in *Heliconius* Butterflies

Opsins are light-sensitive G-protein coupled receptors that determine the wavelengths at which photoreceptor cells absorb light (Briscoe & Chittka, 2001). The ancestral insect visual system is built on three types of opsins: one long-wavelength (LW), one blue (B), and one ultraviolet (UV) (Briscoe & Chittaka, 2001). Butterflies are special as a group that can have more than one LW, B, or UV opsin (Finkbeiner and Briscoe, 2021). These opsins encode photoreceptors that support a variety of tasks, such as mate recognition, foraging, and host plant selection (Finkbeiner and Briscoe, 2021).

Heliconius butterflies can have two UV-sensitive opsin genes, UVRh1 and UVRh2 (Briscoe et al., 2010). Molecular clock and phylogenetic dating studies suggest that the pair of genes arose from a duplication event approximately 18.5 million years ago (Chakraborty et. al, 2023). Both genes are largely conserved across the genus (Heliconius Genome Consortium, 2012). The expression of both UVRh1 and UVRh2 across clades and species is not universal. UVRh1 is absent in *sara* and *erato* clade males (McCulloch et. al, 2017). Moreover, UVRh2 is absent in both *silvaniform* and *melpomene* species, and males and females highly express both

functional copies in the *doris* clade but in different spatial patterns across sexes (McCulloch et al., 2017).

UVRh1 and UVRh2 possess overlapping but distinct absorption spectra, with peak responsiveness at 355 nm and 390 nm, respectively, determined through electrophysiology and microspectrophotometry (McCulloch et al., 2016). *Heliconius* butterflies that retain both UVRh1 and UVRh2 have true UV color discrimination, and in behavioral experiments ran by Finkbeiner and Briscoe (2021), it was determined that *H. erato* females had true color discrimination between 380 and 390 nm (Finkbeiner & Briscoe, 2021). Most other butterflies do not possess this trait, suggesting that there were factors in their environment that deemed both UV opsins necessary to *H. erato* survival.

Behavioral Ecology of UV Vision

Heliconius butterflies protect themselves through Müllerian mimicry, in which multiple similar toxic species converge upon the same warning signals to facilitate predator learning (Kronforst & Papa, 2015). As a result, a paradox is created where butterflies have similar wing patterns and colors to dissuade predators from attacking them, but also need a system to be distinguishable from these same lookalikes when selecting mates. *Heliconius* wings are layered with scales that have *Heliconius* 3-hydroxykynurenine (3-OHK), a yellow pigment that reflects color in the UV range along with fluorescent and LW light. (Finkbeiner et al., 2017). Enhanced UV vision along with UV reflection on wings becomes their solution, allowing individuals of different species to recognize conspecifics while maintaining standardized protective coloration at the community level. This way, the evolution of opsins in *Heliconius* butterflies complements UV signaling, allowing for increased survival.

One of the most important functions of UV wing signals is mate recognition. Behavioral experiments revealed that *H. erato* males selected females that possessed species-specific UV reflection patterns. In contrast, females without these UV cues received significantly less interest (Finkbeiner et al., 2017). This UV-based mate selection is important because of closely related *Heliconius* species living in similar geographic areas. Using UV-specific patterns to filter mates allows males to increase reproduction success.

Unlike most butterflies that only live for weeks in the adult stage, *Heliconius* butterflies not only feed off nectar but also pollen, providing essential amino acids that extend their lifespan up to 6 months (Young & Montgomery, 2020). Since pollen sources are relatively scarce and are usually scattered across forests, it is crucial for *Heliconius* butterflies to be able to detect and recognize these resources efficiently, leading to sophisticated foraging behaviors. Enhanced UV color discrimination likely provides *Heliconius* butterflies with stronger foraging capabilities, such as allowing them to establish “traplines,” foraging routes where specific plants are visited regularly, suggesting that *Heliconius* butterflies navigate through complex environments using visual landmarks (Young & Montgomery, 2020).

Models created by Finkbeiner and Briscoe (2021) have hypothesized that the unique visual system in *Heliconius* butterflies has functional advantages in foraging, especially in

females that require protein for egg production, although further behavioral studies are required for confirmation (Finkbeiner & Briscoe, 2021). Visual studies revealed that when comparing, female *H. erato* (both UVRh1 and UVRh2), male *H. erato* (UVRh2), *H. melpomene* (UVRh1), and *Eueides isabella* (never evolved dual opsins), enhanced UV vision allowed female *H. erato* to distinguish between 380nm and 390 nm reliably, and also discriminate between the UV-reflective yellow inner corollas of *Psiguria* flowers and their outer orange petals, confirming that there is an advantage in having two functional UV opsins (Finkbeiner & Briscoe, 2021).

Comparative perspectives

As discussed above, the typical ancestral butterfly visual system consists of one UV-sensitive opsin, one blue-sensitive opsin, and one long-wavelength-sensitive opsin that corresponds to UV, B, and LW photoreceptors (Briscoe & Chittka, 2001). This composition is extremely common in insect taxa such as *Vanessa cardui* (painted lady), *Danaus plexippus* (monarch butterflies), the moth *Manduca sexta* (tobacco hornworm), the bees *Apis mellifera* (honeybee) and *Bombus impatiens* (bumblebee) (Wakakuwa et al., 2010). In *Manduca sexta*, for instance, three rhodopsins with peak sensitivities at 520 nm (green), 450 nm (blue), and 357 nm (UV) provide the standard trichromatic foundation for color vision (White et al., 2003). Furthermore, *Danaus plexippus* have one UV opsin that collaborates with blue receptors to control navigation and foraging (Stalleicken et al., 2006). However, because such species do not possess multiple UV receptors, they can not perform true UV color discrimination. Instead, they depend upon contrasts between UV and other wavelengths.

Papilio xuthus (swallowtail) has six photoreceptor classes as a result of opsin duplications through several ranges of light (Koshitaka et al., 2008). However, even then, it evolved only true color vision across blue and long-wavelength receptors, with only a single UV opsin. It suggests that the evolution of opsins is based on lineage, shaped by an organism's environment instead of following along a single evolutionary trajectory, highlighting how unique *Heliconius* butterflies are, having a visual system that arose from lineage-specific innovation that solved ecological challenges, leading to specialized foraging and protection systems.

Conclusions and Future Directions

The *Heliconius* genus is extremely indicative of how gene duplication can cascade across different biological levels. At the molecular level, opsin duplication resulted in increased color sensitivity. At the physiological level, it allowed *Heliconius erato* females to develop true UV color discrimination. At the behavioral level, it enabled ecological strategies such as pollen feeding and sex-specific oviposition site selection. Most importantly, at the evolutionary scale, gene duplications enable the reinforcement of Müllerian mimicry and diversification of color vision capabilities.

Despite significant research, there are still important gaps. Questions for future research include: How do sensory innovations like enhanced color vision drive ecological specialization? What role does sexual dimorphism in sensory systems play in promoting reproductive isolation

and speciation? How do complex mimicry systems shape the evolution of communication channels?

Heliconius butterflies bridge genomics, physiology, and behavior, an extraordinary model for sensory-behavioral coevolution. As environmental pressures reshape ecosystems, insight from studying *Heliconius* butterflies could provide crucial information for predicting how other species with specialized sensory adaptations will respond to changing conditions.

Works Cited

- Briscoe, Adriana D., et al. "Positive Selection of a Duplicated UV-Sensitive Visual Pigment Coincides with Wing Pigment Evolution in *Heliconius* Butterflies." *Proceedings of the National Academy of Sciences*, vol. 107, no. 8 (2010): 3628-3633. Accessed September 9, 2025. <https://www.pnas.org/content/107/8/3628>.
- Briscoe, Adriana D. and Chittka, Lars. "The Evolution of Color Vision in Insects." *Annual Review of Entomology*, vol. 46 (2001): 471-510. Accessed 9 Sept. 2025. <https://www.annualreviews.org/content/journals/10.1146/annurev.ento.46.1.471>.
- Briscoe, Adriana D. "Reconstructing the Ancestral Butterfly Eye: Focus on the Opsins." *Journal of Experimental Biology*, vol. 211, no. 11 (2008): 1805-1813. Accessed 21 Sept. 2025. <https://journals.biologists.com/jeb/article/211/11/1805/9503/Reconstructing-the-ancestral-butterfly-eye-focus>.
- Chakraborty, Mahul, et al. "Sex-linked Gene Traffic Underlies the Acquisition of Sexually Dimorphic UV Color Vision in *Heliconius* Butterflies." *Proceedings of the National Academy of Sciences*, vol. 120, no. 33 (2023): e2301411120. Accessed September 19, 2025. <https://www.pnas.org/doi/10.1073/pnas.2301411120>
- Finkbeiner, Susan D. and Briscoe, Adriana D. "True UV Color Vision in a Female Butterfly with Two UV Opsins." *Journal of Experimental Biology*, vol. 224, no. 11 (2021), article jeb242064. Accessed 15 Sept. 2025. <https://journals.biologists.com/jeb/article/224/11/jeb242064/272299>.
- Finkbeiner, Susan D., et al. "Ultraviolet and Yellow Reflectance but Not Fluorescence is Important for Visual Discrimination of Conspecifics by *Heliconius erato*." *Journal of Experimental Biology*, vol. 220, no. 7 (2017): 1267-1276. Accessed 15 Sept. 2025. <https://journals.biologists.com/jeb/article/220/7/1267/46832>.
- Gerl, Ellen J. and Morris, Molly R. "The Causes and Consequences of Color Vision." *Evolution: Education and Outreach*, vol. 1, no. 4 (2008): 476-486. Accessed 7 Sept. 2025. <https://evolution-outreach.biomedcentral.com/articles/10.1007/s12052-008-0088-x>.
- The *Heliconius* Genome Consortium. "Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species." *Nature*, vol. 487, no. 7405 (2012): 94-98. Accessed 10 Sept. 2025. <https://www.nature.com/articles/nature11041>.
- Koshitaka, Hisaharu, et al. "Tetrachromacy in a Butterfly that Has Eight Varieties of Spectral Receptors." *Proceedings of the Royal Society B*, vol. 275, no. 1637 (2008): 947-954. Accessed 7 Sept. 2025. <https://royalsocietypublishing.org/doi/10.1098/rspb.2007.1614>.
- Kronforst, Marcus R. and Papa, Riccardo. "The Functional Basis of Wing Patterning in *Heliconius* Butterflies: The Molecules behind Mimicry." *Genetics*, vol. 200, no. 1 (2015): 1-19. Accessed 10 Sept. 2025. <https://academic.oup.com/genetics/article/200/1/1/5936100>.
- McCulloch, Kyle J., et al. "Sexual Dimorphism and Retinal Mosaic Diversification Following the Evolution of a Violet Receptor in Butterflies." *Molecular Biology and Evolution*, vol.

- 34, no. 9 (2017): 2271-2284. Accessed 10 Sept. 2025.
<https://academic.oup.com/mbe/article/34/9/2271/3829953>.
- McCulloch, Kyle J., et al. "Sexual Dimorphism in the Compound Eye of *Heliconius erato*: A Nymphalid Butterfly with at Least Five Spectral Classes of Photoreceptor." *Journal of Experimental Biology*, vol. 219, no. 15 (2016): 2377-2387. Accessed 10 Sept. 2025.
<https://journals.biologists.com/jeb/article/219/15/2377/58418>.
- Stalleicken, Julia, et al. "Physiological Characterization of the Compound Eye in Monarch Butterflies with Focus on the Dorsal Rim Area." *Journal of Comparative Physiology A*, vol. 192, no. 3 (2006): 321-331. Accessed 7 Sept. 2025.
<https://link.springer.com/article/10.1007/s00359-005-0073-6>.
- Wakakuwa, Motohiro, et al. "Evolution and Mechanism of Spectral Tuning of Blue-Absorbing Visual Pigments in Butterflies." *Naturwissenschaften*, vol. 97, no. 2 (2010): 137-144. Accessed 7 Sept. 2025. <https://doi.org/10.1371/journal.pone.0015015>.
- White, Richard H., et al. "The Retina of *Manduca sexta*: Rhodopsin Expression, the Mosaic of Green-, Blue- and UV-Sensitive Photoreceptors, and Regional Specialization." *Journal of Experimental Biology*, vol. 206, no. 19 (2003): 3337-3348. Accessed 7 Sept. 2025.
<https://journals.biologists.com/jeb/article/206/19/3337/14017>.
- Young, Fletcher J. and Montgomery, Stephen H. "Pollen Feeding in *Heliconius* Butterflies: The Singular Evolution of an Adaptive Suite." *Proceedings of the Royal Society B*, vol. 287, no. 1938 (2020), article 20201304. Accessed 15 Sept. 2025.
<https://royalsocietypublishing.org/doi/10.1098/rspb.2020.1304>.

Whether Corporate Governance in India Post Liberalization and Colonialism Shows No Distinction from Pre Liberalization and Colonialism By Trayambika Reddy Chirra

Abstract

Corporate governance has only changed to benefit the government and its subjects, despite substantial efforts in reforming corporate governance in India, it is still a reflection of corporate governance during colonialism. In this paper, I will prove that corporate governance in India post liberalization and colonialism shows no distinction from pre liberalization and colonialism. The BharatPe case (2022) and Haridas Mundhra Scandal (1957) serves as a testament to the claim that corporate governance challenges persist across economic eras regardless of law reforms. It's not just how seriously the corporate sector in India perceives the law but also how effective the law and the governance is within the system. I will identify the current corporate governance regime within its trajectory of historical development from the early extractive corporation form developed by the East India Company to drain wealth out of the country. This model will be used to illustrate the contrasting themes with the current corporate governance. I will argue that the change is non-existential and superficial, since it failed to erase the corporate structures habituated in the colonial era. System change can be specifically referred to the fundamental shift in the governmental schemes, accountability mechanisms, and power dynamics within corporate governance frameworks. Even with legislative reforms like Companies Act (2013) and liberalization of India (1991), the grounds for change have not been updated, since they have not effectively altered the underlying patterns of corporate misconduct, elite economic control and irregular government schemes. I will highlight the major errors in the construction of the corporate governance and laws, and question whether corporate productivity is possible without a paradigm shift. In essence, with the observations and evidence suggested in the paper, I conclude that corporate governance in the present day is similar to corporate governance structure during colonial rule, proving that to anticipate change we have to alter the dictatorial methods of India.

Introduction

Corporate governance refers to the legal framework built on the basis of rules, practices, and processes by which a company is directed and controlled. This system ensures accountability, transparency, and fairness to protect the interests of stakeholders, including society, shareholders, management and others⁶⁹. Corporate governance manifests a recovery from an unjust society to a profitable yet fair political system. Although this system aims to reach a fair political system, to what extent it is efficient can be questioned, since this framework has numerous layers that challenge its overall goal. To achieve transparency, fairness, and accountability in a way that it accounts for the economic and social needs. Specifically, on the forefront, it seems as if corporate governance in India has changed over the years, but from a

⁶⁹ **Wadhwa, Suhani.** "Corporate Governance: A Legal Framework for Transparent and Ethical Business." *LegalServiceIndia*. Accessed 31 Aug. 2025.

nuanced perspective, it can be identified that there is so much it lacks. I hypothesize that there has been no concrete change in corporate governance in independent India and that despite reforms the system remains similar to colonial era structures. This is evident through rising cases like the Haridas Mundhra Scandal (1957) and the BharatPe controversy. Legislative changes are highly superficial because the Companies Act (2013) and economic liberalization (1991) have not fundamentally altered corporate misconduct and elite control. This elite control could be the primary issue, since government organizations become weak as control is maintained by the upper class or the affluent people in the society. This showcases how corporate governance today replicates the extractive model of the East India Company, evading equitable economic control. This summarizes the problem, that government schemes and power dynamics are flawed, as a result, there is a requirement for a paradigm shift that encompasses corporate productivity and genuine reforms, rather than superficial legal updates.

In India, there are several legal mechanisms that are meant to ensure corporate governance; however their effectiveness is undermined by weak enforcement, elite control, and systematic loopholes. To undergo structural amendments these legal mechanisms must be changed to fit the new shift. The mechanisms at stake can be dealt with in several corporate sectors, including corporate law, accountability mechanisms, judicial oversight and economic control. Explicitly, within the boundaries of corporate law Companies Act, which was established in 2013 by the parliament of India as a primary source of Indian company law, must be fixed. Similarly, Securities and India Board Regulations (SEBI), the Insolvency and Bankruptcy Code (IBC) 2016, and Foreign Exchange Management Act (FEMA) 1999, all need to be altered to anticipate change. Specifically, it is not the law that needs to change, but the government and the elite control is what needs to change to a great extent. The rise in cases of corporate mismanagement and fraud can be attributed to legal evasion and the manipulation of regulatory frameworks, often facilitated by the privileged status of wealthy institutions and their elite connections. Accountability mechanisms in India have a weak control over corporate governance, due to constantly being dismissed from mandating accountability of corporate misconduct. This reduces their role in ensuring a just system. Including board of directors and independent directors that are essentially required for mandating fair corporate governance, but due to political and elite business influences their role completely changes into a superficial regulatory body⁷⁰. Moreover, within Auditing and financial reporting laws, and Corporate Social Responsibility Mandate (CSR)⁷¹, it has been revealed through fraud cases that these mechanisms have regulatory failures, misuse of funds, and manipulated financial reports⁷². All of these issues lead back to elite and governmental influence and the barrier between a just and an unjust corporate system.

To dismantle the current corporate structure and identify all the issues at stake, the initial reformations in Indian company law have to be scrutinized. Within the abundance of corporate layers liberalization plays an integral role in this system. The liberalization of the Indian

⁷⁰ "CAG Flags Governance Issues among Central Public Sector Enterprises." *The Economic Times*, 24 Feb. 2025.

⁷¹ "A Crisis of Confidence in Indian Auditing." *The Economic Times CFO*, 12 Feb. 2025.

⁷² "I-T Raids in 6 States to Uncover CSR Fund Diversion Scam." *The Times of India*, 10 Mar. 2025.

economy (1991) marked a fundamental shift from a state-controlled, protectionist and command economy to a market-driven globalized economy. Present-day India is an example of a mixed economy where both market and governmental forces drive the corporate system, which is due to liberalization. Liberalization was triggered by a severe balance of payments crisis, this led to India seeking financial assistance from the International Monetary Fund (IMF). As compensation the IMF mandated that India shift their corporate structure, including reducing government control over industries, to encourage foreign investment, enhancing Indian market presence in the global economy, and further emphasizing on globalization. The liberalization reforms dismantled the license Raj, which essentially was a scheme that dictated who established companies in India. By removing license Raj from the Indian economic equation, there was a reduction of trade barriers which opened up various sectors to big foreign and private players. This transformation repurposed the economic and corporate landscape of India.

With the growing number of companies in India, the government had to establish an abundance of legal frameworks that could possibly tackle the rise of crony capitalism and corporate scandals. While liberalization promoted a competitive market, it also led to increased corporate-political nexus, where businesses aligned with political power to achieve elite control and advantages. An example of this situation is the accusation that Reliance industry is benefiting from political connections by tailoring policies. Dhirubai Ambani's Reliance has also seen steadfast growth in the market pre-liberalisation, the structural change gave them an opportunity to expand. Due to its close alliance with the government including Congress and BJP, they were accused of manipulating and controlling policy decisions in the country⁷³. This included petrochemical policies and tariff changes in Reliance's favor⁷⁴, which sparked attention on the idea that elite influence can be evident regardless of structural changes to corporate governance. Regarding the Krishna Godavari (KG) Basin gas pricing controversy, Dhiru Ambani was accused of using political connections to get support from the UPA government to benefit from inflated gas prices. This favor system reflects the themes of license Raj as in both scenarios the government chooses businesses to benefit each other⁷⁵. Regardless, the expansion of businesses through liberalisation has led to increased global attention on India, GDP growth, job opportunities, and demand for innovation. To block the rise of increased social welfare loss, the government introduced new corporate laws and reforms like The Companies Act 2013, SEBI, Competition Act 2002, and FEMA. Despite persistent efforts, the government lacks the capability of maintaining the balance resulting from weak implementation. Liberalization is a prime example of the larger step India has taken to escape from financial troubles and incentivise a growing market, it is a pivotal point in corporate governance. It marked the shift from a command economy to a mixed economy, but failed to democratize corporate power, allowing further emphasis on elite groups that control wealth. Liberalization can be contrasted with

⁷³ "The Times of India." "Reliance Favourite Group of BJP and Congress, Kejriwal Says." *The Times of India*, accessed 31 Aug. 2025.

⁷⁴ *Voice of Reliance*. Frontline, The Hindu Group. Accessed 31 Aug. 2025.

⁷⁵ Guha Thakurta, Paranjay, Subir Ghosh, and Jyotirmoy Chaudhuri. *Gas Wars: Crony Capitalism and the Ambanis*. Paranjay Guha Thakurta, 2014.

colonial and present day corporate governance to prove that these changes were illusions rooted to colonial extractive policies, which will remain superficial until deeper systematic changes prosper.

Further moving back in time, colonial rule is a root layout of the corporate structure of India present day. It could be hypothesized that India still follows the traditional legal patterns from the colonial era. This structure dismisses fair and equal choice, additionally mandating citizens to follow a dictatorial regime. The legal system during the colonial rule leaned on authoritative rule used to manipulate citizens, specifically the corporate structure. Most companies established during that time were permitted to perform business functions, since they supported the British Raj, however this permit system was only valid to those supporting the colonial rule. Tata Sons has obtained this privilege, they gained a degree of favor from the British Raj, which helped them scale up operations, and become what it's known today, through loyalty to the colonial economic system. The British Raj found Tata to be beneficial to their stance in India, since it had a non-confrontational stance and industry driven goals. It was noted that during World War 1, Tata Iron and Steel Company which was set up in 1907 received preferential treatment from Britain, since they sourced steel for the war effort. This permit system is similar to the pre-liberalization corporate structure, since it followed a command economy pattern similar to the colonial era. Another pivotal evidence of the superstitious change is the elite control, although present day India doesn't admit to elite control, it is inherently embedded in legal and governmental systems.

Reaffirming the hypothesis that Indian corporate governance is similar to the pre-liberalization and colonial models. If significant reforms and legal advancements haven't changed the mirrored reality of corporate governance to colonial era and pre liberalization, the society will only anticipate corporate misconduct rather than effective reforms that erase elite control and irregular government hierarchy. Moreover the historical trajectory of corporate governance in India can be a background that India needs to move away from to foster the four pillars of corporate governance - transparency, accountability, fairness and responsibility.

Section 1

Colonial rule in India was a period in India's history that substantially changed governance and most importantly the 'people's world'. A period of 200 years in India's history was inflicted by dictatorship and British rule, and this era was a time where India's rich culture and values started altering. From 1757 to 1935 British control from the East India company and later under direct crown rule drained the wealth of the country through policies that aligned with their interests. Additionally several other schemes were implemented including exploitative land revenue systems (Permanent Settlement, Ryotwari, Mahalwari), and the destruction of indigenous industries. These changes to the Indian governance system affected the corporate structure as it was expensive for Indian citizens to establish their wealth and riches without being exploited by the government, simultaneously they had the obligation to support British governance to continue their business functions. The government structure can be defined as

authoritative and hierarchical, which established a bureaucratic structure that persists today. This similarity showcases how the interests haven't changed, and the government has to deal with these loopholes to correct market failures. First-hand view on this issue can be deceptive, as people delve into the models referred to by the government, we start to realize that it mirrors the colonial and pre-liberalization structure. This questions whether laws and policies can truly change and refine the loopholes and existential corporate models.

William Dalrymple's book 'Anarchy' provides a detailed account of the relentless rise of the East India Company. The author charts the disintegration of India into a state of civil strife and atomisation, which provides a valuable window of entry through which the British sealed their rule over the sub-continent. This book shows how the East India Company, a private corporation, transitioned from holding a position within Indian society to ruling India. This specifically brings into picture the embedded corporate structure in colonialism. The book argues that the East India Company's motive was to initially establish a corporate dictatorship through military force, political manipulation, elite control, and economic exploitation. This scheme slowly spread to other vast territories, which can be symbolized as a 'disease'. Dalrymple's insight into the world of colonialism is significantly attached to corporate governance in India, illustrating how unregulated corporate power influenced by political influence can shape economies and societies. The book provided a multifaceted account of the British rule over India, which served as a basis for acknowledgement for the elite control present throughout times. The book wasn't meant to retell the same story or present a new view of the knowledge of colonialism, rather to present the story from the view of Mughal nobles, English merchants, and Indian financiers. This analysis of the book 'Anarchy' establishes the foreground of where to look when proving the hypothesis that the elite control which was apparently dismantled after the colonial era, still persists today.

Solomon vs Solomon is a case that illustrates how cases can bring new perspectives to the legal system and change how the economic system runs. Particularly, this case presented the landmark judgment in company law regarding the principle of corporate personality and the lifting of the corporate veil. It had profound implications for the development and reformation of the corporate system in England. To summarise, the House of Lords ruled that the company was a separate legal entity from its owner, and although the owner controlled it entirely. This case was the basis for confirmation of the principle of corporate personality, it promoted limited liability, boosted entrepreneurial activity, and triggered legal and policy reforms. The case promoted several motives including encouraging economic growth, facilitating access to capital as companies were able to easily raise funds through raise insurance. Additionally, it provided legal certainty as it increased predictability in business dealings, and supported efficient industrialisation. A similar case contrasting to the Indian economic situation is *Bacha F. Guzdar v Commissioner of Income Tax, Bombay* (1955). It questioned whether or not a shareholder has a right in the company's income or assets in the same way the company does? Similar to the case in England, this case also ruled that a company is a separate legal person, distinct from its shareholders, and that the income of the company should not be considered as the income of the

shareholder. Both cases have introduced a new legal perspective; however, the replication of the laws produced through *Solomon v Solomon* through the Companies Act, 1956, resulted in a different pathway in the Indian economy. This legal structure set precedent for shareholder rights, influenced tax law interpretation, and confirmed corporate personality. Simultaneously, it increased the chances of misuse as it led to the creation of shell companies to evade tax or hide assets, and using the corporate structure to commit fraud with limited personal consequences. To combat this, the Indian courts evolved doctrines like lifting or piercing the corporate veil.

A paramount number of putative rule concepts of corporate law that were devised during the colonial period can be identified. Modern corporate governance, limited liability and corporate personhood are foundations that were shaped by the East India Company (EIC) and other colonial-era corporations. Some of the important principles that can be contrasted with present-day foundations are fiduciary responsibility and governance, limited liability, chartered corporations and monopoly rights, separation of ownership and management, and corporate accountability. Although these policies were administered to cater to British rule, it served as a basis for further establishment of this knowledge. These ideals were reformed and re-established in the Companies Act, 2013 and SEBI regulations. This shows how subjectivity in this case can be later transformed into collective objectivity. During both periods of Indian corporate governance, hierarchical influence is present, possibly due to the lack of separation from colonial strategies. Unconsciously as the government and the law link their methods to colonial rule, they adopt erroneous techniques that could potentially fail India's bureaucratic corporate structure, regulatory frameworks, and governance.

The British governance corporatised and legally appropriated various customary practices including the Hindu Undivided Family (HUF). This development legally allowed recognition of joint family properties, as it was reinterpreted to function as a corporate entity. This changes the corporate structure during the colonial rule as family owned properties were recognized as a separate legal entity distinct from its owners. This knowledge can be linked with the 'corporate personality' law that was reintroduced into corporate structure after liberalisation, where the entity is given rights, liabilities and tax benefits. Ultimately, it erased the fine line between family wealth and corporate structures. The British rule essentially codified the HUF structure into legal terms making it a judicial entity that is separate from the stakeholders controlling it. (More depth on the link between HUF and Income Tax Act) Additionally, according to the British legal system this change allowed for optimal taxation that was later adopted by India in the authorisation of Income Tax Act, 1961.

The opium trade and the rise of early corporations can be interlinked with the period of British colonialism. It could be responsible for the expansion of global capitalism in the 17th to 19th centuries. This situation reveals how early corporate power was built on exploitation, elite control, colonial violence, and extractive trade systems. Before and during the late 18th century, Britain had a huge trade deficit with China because of high demand for Chinese silk, tea, and porcelain. To resist further implications, the British East India Company began illegally exporting opium grown in India to China, which led to widespread addiction and health

concerns. Bihar and Bengal were largely controlled by EIC in exploiting them of their resources to offset the trade deficit. This opium trade had lasting negative externalities on China, since it resulted in millions of Chinese being addicted to Opium. Ultimately, this led to the rise in black markets for opium smuggling and dependence on the informal sector of the economy increased in China. In essence, Britain was gaining power and profits through the suffering of China. Britain actively disregarded the morality in this situation, which led to harm to both parties-India and China. To counteract the Chinese government's start seizing warehouses holding opium in China, these consecutive events are the primary reason for the opium wars between China and Britain. This case clearly exemplifies the typical nature of colonialism and how it obstructed justice and morality.

From a detached perspective it may seem that India has long moved away from colonial patterns⁷⁶, however, this is just a facade, since within our systems we still engage in crony and corrupt strategies⁷⁷ due to elite influence. Although reforms like economic liberalisation and companies act, 2013 have been established the laws in India are ineffective if these loopholes in management are not eradicated. There are laws present in India that are a colonial carryover, and they could be the reason for the constant mismanagement of the system. For example, the sedition law is a classic example of this colonial carryover, since, during British rule, sedition law was imposed to control and subdue the power of citizens. This has not been eradicated altogether by the current Indian government, which demeans the power of word of the citizens of India, and their ability to raise questions against the ruling party in fear of arrest or other forms of suppression. This oppressive nature of a few laws could be the reason why there are several frauds and crimes. The crony business and corporate nexus today is a reflection of the same forces that streamlined colonial-era exploitive systems like the opium trade, this contrasting situation highlights the persisting weak accountability and political influence. Crony capitalism is often seen in India, since there is a close relationship between business ventures and government organisations. This stimulates bias and unfair competition which may cause regulatory capture, monopolies, oligopolies, and corruption. These connections also exist in electoral bonds, media ownership, and infrastructure contracts, overall this dictates what is knowledge in a society.

Cases that reveal the mismanagement of corporate governance are the BharatPe case⁷⁸ and the Haridas Mundhra Scandal⁷⁹. Haridas Mundhra, a Kolkata based industrialist, exploited the government owned life insurance company (LIC) into investing Rs.1.26 crore to buy shares in his financially struggling companies. In 1957, Rs.1.26 crore was a large sum, and this led to numerous financial discrepancies in the market. The value of the shares were overvalued, therefore this led to scandal break by Feroze Gandhi and an Indian Express journalist Ramnath Goenka. Simultaneously, there were several judicial inquiries and Mundhra was found guilty and

⁷⁶ Tully, Mark. "Governance in India Is a Relic of Its Colonial Past." *Hindustan Times*, 27 Jan. 2019. Accessed 31 Aug. 2025.

⁷⁷ *Crony Capitalism and State Capture*. *Frontline*, The Hindu Group. Accessed 31 Aug. 2025.

⁷⁸ Mishra, Digbijay. "BharatPe Removes Ashneer Grover from All Positions after Governance Review." *The Economic Times*, 2 Mar. 2022.

⁷⁹ Ramakrishnan, T. "Haridas Mundhra Case, 1957: India's First Financial Scandal." *The Hindu*, 18 Dec. 2019.

sentenced to prison. This corporate-political scandal highlights the inefficiencies in the system, and set precedent for ministerial responsibility. A recent case that illustrates a similar problem is the BharatPe case, where the fintech startup which is the part of India's startup boom, was noticed for fraud. Internal audits accused Ashneer Grover the founder of lavish personal spending of company funds, and using fake invoices to siphoning money. This led to the filing of the lawsuit and simultaneously the removal of Ashneer Grover and his family from company positions. Both these cases show how mismanagement still exists, weak institutional accountability, and how personal power is used to bypass checks.

To summarise the economic landscape of India throughout the years- before colonisation India was a indigenous community composed of agriculture reliant industries, and small decentralised markets. During the colonial era British control increased, which introduced extractive practices, corporate monopolies, and political cronyism. After the end of the British Raj, there was a replacement by a highly controlled economic environment that was somewhat similar to the economic structure during the British Raj. Due to this crony capitalism emerged, with large industrial corporations like Ambanis, Birlas, Tatas. These companies received continuous favours from the government, often through government investment into loss-making companies. This reflects the extractive practices during the colonial era, as companies started bribing to receive favours from the government. The structure of the relationship between corporations and governments remains similar to the colonial era, due to this system law became redundant.

Section 2

Pre-liberalisation to liberalisation emphasizes the shift from a command economy to a market-based economy⁸⁰. During pre-liberalisation, the government controlled what was produced and sold in the market, however with the change the government's control was minimised. The liberalised economy demonstrates more involvement and advancement in entrepreneurial growth, since the government did not control the production of goods and services. Although, India is not an example of a stage 6 economy or a complete market based economy like the U.S, it is identified that it is driving itself to a market based economy from a mixed economy where both government and private corporations are involved in business production. The Immediate pre-liberalisation stage was composed of a strict License Raj, that promoted monopolies in the economy. The economic activity involved state-led industrialisation, License Raj, Import substitution, and centralised government planning. This led to economic stagnation by 1981, during this phase India was facing several economic issues with growth and employment. Their main issue was economic stagnation, high inflation which raised prices, persistent poverty, due to high inflation and poverty, the government couldn't sustainably provide resources to citizens⁸¹. The 1991 Liberalisation was primarily due to these increasing issues, and

⁸⁰ "Independence Day 2025: Tryst with Growth — India's Economic Journey from Nehru to Now." *The Economic Times*, 15 Aug. 2025.

⁸¹ CSR Education. *The 1991 Economic Reforms: How India Transitioned to a Market-Oriented Economy*. May 10, 2024. *Development in India*, CSR Education, 2025. Accessed 31 Aug. 2025.

in attempts to decrease unemployment and increase economic growth, India shifted to a liberalised economy.

This major economic reform shifted India into a more market-oriented economy, prioritized economic development, and opened its arms to international firms which increased education and job levels in India. India today is as a result of this liberalisation and the recovery from colonial times. As international firms started establishing businesses in India it sparked innovation and education as the government started installing proper education systems that sponsored the new generations. In specific, the liberalisation devalued the rupee to address the balance of payments crisis, this ultimately made exports more competitive⁸². Additionally, it introduced policy reforms that deregulated industries by reducing the need or licenses (dismantling the License Raj). Further on, it also helped Indian companies to expand and diversify without bureaucratic blocks. The trade policy was also altered to fit the expanding economy, for example tariffs imposition was more loose, which reduced import restrictions. This allowed for economic boom, as foreign direct investment (FDI) allowed for local businesses to grow and become profitable⁸³. Privatisation instead of tight state controlled enterprises (SOEs) and public sector undertakings (PSUs) offered improved efficiency and reduced fiscal burden. With the financial sector reforms, the Reserve Bank of India changed its mechanisms and adopted more market-based mechanisms to control inflation, interest rate, and exchange rate. All of these positive externalities of this economic shift led to eradicating the scarce minimum per capita GDP, while also increasing global recognition⁸⁴. Simultaneously, the burden on the government also decreased since pre-liberalisation dealt with numerous issues with decreasing the reliance on the informal sector, however, with this change the government can decrease their focus on social welfare loss.⁸⁵

Corporate governance was heavily influenced due to these apparent changes, although it dismantled oligopolies and monopolies in the market, by increasing greater competition, and economic efficiency, it scrutinized the legality of the functions. Stronger regulations and governance frameworks needed to be introduced to accommodate the changes, since the old policies wouldn't completely account for the change in corporate thought. Since pre-liberalisation times were heavily attached to the colonial era, and was a period for recovery from the extractive model of colonialism, the post-liberalisation needed a change in the system for transparency and fairness. There were an increased number of loopholes in the system and attention in legal functions was more necessary than ever. Due to this change there was an imbalance between order and stability as legal functions did not accommodate the new structural improvement. The government introduced organisations like the Securities and Exchange Board of India (SEBI) which essentially regulated stock markets and protected investor interests. This also introduced mandatory disclosure norms, insider trading regulations, and requirements for

⁸² Cerra, Valerie, and Sweta Chaman Saxena. "What Caused the 1991 Currency Crisis in India?" *IMF Staff Papers*, vol. 49, no. 3, 2000, pp. 395–425. *International Monetary Fund*. Accessed 31 Aug. 2025.

⁸³ TEAM. "India at 75: Epochal Moments from the 1990s." *Frontline*, 18 Aug. 2022. *The Hindu Group*. Accessed 31 Aug. 2025.

⁸⁴ The World Bank. "India: Overview." *World Bank*, 16 Sept. 2024. Accessed 31 Aug. 2025.

⁸⁵ Centre for the Advanced Study of India. "The Evolution of India's Social Welfare Regime and Future Challenges." *Center for the Advanced Study of India*, University of Pennsylvania, n.d. Accessed 31 Aug. 2025.

independent directors to tie up the loose ends of decreased government intervention. Liberalisation ensured that the corporate sector of India had more freedom but had to ensure that they were bound by regulations internally, this is why organisations like SEBI were introduced. Along with SEBI companies were now required to disclose certain information regarding finances, follow accounting standards, and several other stock exchange policies to guarantee transparency in corporate dealings. Liberalisation was an unfamiliar concept to the Indian economy, since it was largely practiced in the west, due to this they had to adopt international norms. These standards were particularly practised in developed countries like the United States, United Kingdom, and other EU members. Globalisation efforts started commencing as the Indian government sought to compete with international companies and trade skills and specialisations. To push reforms several committees were also introduced like the Kumar Mangalam Birla committee (1999) and Naresh Chandra committee (2002). These standards allowed leverage even when independence was given to business corporations.

Even with these standards the scams and irregularities in corporate governance persisted; it was a matter of time before the corporate sector of India collapsed due to illegal means of business activities. Satyam scam (2009), Global Trust Bank scam (2004) and other scams related to stock markets displayed how the standards initially set after the liberalization were outdated which failed to keep up with globalisation efforts and the modern structures of the global business environment. To correct market failure the Indian government introduced the Companies Act 2013. This act was crucial since it not only accommodated the changing global business environment, but also urged the anthem of corporate governance which is transparency, accountability, and investor protection. Simultaneously it ensured that there was a greater influx of foreign investments to align with global best practices. It simplified business processes as it shifted the procedural and bureaucratic business processes to a more healthy environment that encourages entrepreneurship and supports startups and businesses. Key features that allowed greater business discretion and legal adherence were included in this Act.

India was the first country to mandate Corporate Social Responsibility (CSR) which means companies meeting certain thresholds must invest 2% average net profits on CSR activities. The relevant provisions that mandate CSR activities is section 135 of companies Act, 2013 and Schedule VII discusses the list of activities that can be considered under CSR. This affirmed correction of social welfare loss and spillover of negative production and consumption externalities. Specifically in corporate governance this act mandated companies to have at least $\frac{1}{3}$ independent directors, introduced board committees and performance evaluation functions. Inspection of business activities have been further scrutinised as stricter auditing reforms have been made which specifies that there is mandatory rotation of auditors for every 5 -10 years and any negligence and fraud can make the auditor criminally liable. 2008 to 2014 was considered as the Startup growth years and dates following that would be considered as the Startup boom phase. These activities were primarily due to the simplification of processes for private and small companies, the Companies Act introduced the concept of One Person Company (OPC) catered towards individual entrepreneurs; another concept is the small company which is for ease of

compliance and reporting. These concepts gradually changed the Indian economy since people did not view business as a far off world which was complicated and had no place for low to middle-class citizens. In a way it eradicated the term class as it allowed people to expand their thought and income. Haridhas Mundhra scandal was another case that dealt with investor fraud, since during the 1950s his unsupervised stock exchange account accumulated funds from people across India and shook India's market and government. This exposed the nexus between the bureaucracy, stock market speculators and small rogue businessmen. History should not repeat itself since a scam of this magnitude did cause a lot of economic chaos, the Companies Act considered these scams and introduced investor protectionist reforms. The Serious Fraud Investigation Office (SFIO) and class action suits were components that corrected this nexus. Overall, stricter penalties, emphasis on disclosure and transparency, and new-age reforms like technology use have been made to safely guide Indian corporate governance to a better place.

Despite the new Act, the scams continued, and people continued to find loopholes in the system. This is not just a question of lacking adequate laws but an enduring cultural and institutional ecosystem that has been going on since the times of colonial extraction, and even today is built on more misgovernance than governance. If you take a scam like the Satyam scam (2009) and question whether it would have stood a chance against Companies Act 2013, the answer would be no. If you compare the situation before and after the introduction to Companies Act 2013, it becomes evident that the legal and governance was significantly strengthened to combat major issues. I can acknowledge that liberalization and companies Act have made imperative changes to the Indian economy that corrected law and order but certain loop holes and government schemes easily seep its way if not addressed appropriately.

Conclusion

To summarise, within the boundaries of law, corporate governance portrays itself as a prominent topic that establishes a foreground for corporate companies to guide themselves. Corporate governance has developed incessantly to refine the question "To what extent should a company's managers be held responsible by law?". This development is as a result of certain cases that have interpreted the knowledge of corporate governance differently, and have raised questions through court. Cases like *Salomon v A Salomon & Co Ltd*, *The Citizen's Life Assurance Company v Brown*, and *Foss v Harbottle*. The corporate bodies have been identified with active manipulation of the law through loopholes in corporate governance. Through these cases, courts have coined terms like corporate personality, and corporate veil, which allows the company to be recognized as an individual entity. Companies like these, have been criminally charged for practicing malicious acts under the law scheme, this allowed courts to refine existential corporate law.

A filtered version of the corporate law still doesn't solve all issues, since certain cases allow courts to rectify loopholes within corporate governance to this day. The BharatPe case, is also one that falls under the radar of corporate governance like all corporate bodies, although specifically dealing with many allegations against parties within the company management. This

demonstrated how laws of corporate governance are projected on individual corporate cases. Similarly, in this case BharatPe had filed a criminal complaint in December 2022 against Mr. Grover, who was the co-founder and former managing director of BharatPe. This is to inform Mr. Grover and his family faced allegations of misappropriating company funds amounting to rs.81 crore. The Economics Offences Winger (EOW) of the Delhi Police filed a FIR for the claims of siphoning off funds from BharatPe through various fraudulent activities. There are several key allegations against Mr. Grover and his Family, appealed by BharatPe shareholders, including fake invoices for money transfer, misappropriation of funds, and related entities.

Although under the law, it is an assumption that illegal activities undertaken by the company without being legally responsible for the social welfare loss, due to laws like corporate personality and corporate veil. These are specific laws under the companies Act, to elaborate on the safety and privacy of the shareholders of the company. Regardless of this protection, it was ruled that corporations can also be held accountable for their actions indicating malicious intent, or violation of company laws. In BharatPe's situation, there were several violations involving Ashneer Grover, involving financial misconduct, lack of transparency, weak internal controls, and board oversight failures. This principle was evident in the BharatPe case, therefore the legal proceedings suggested a settlement, withdrawal of legal actions, quashing of FIR and lookout circular. Subsequently, after the settlement, Grover was disassociated from the company entirely, relinquishing any management roles, and Delhi High Court quashed the FIR against his wife.

For corporate governance to improve In India, different organs of state need to improve simultaneously. From my readings I was able to interpret that Indian law is as complex and functional as it is, but for it to be actionable and for it to make an impact in justice and order, structural changes in hiring political bodies should be changed, enforcement laws should be changed, and abuse of hierarchy needs to be dismantled. We have been following certain laws like the sedition law under the comfort of the government, but we must eradicate this to establish freedom of speech and dismantle these power and money influences. This can be achieved through merit based appointments for political bodies, since often parties elect bodies based on finance, income and power. Additionally we must plan on strengthening enforcement laws such as SEBI, CCI and MCA, whistleblower protection in public and private sectors, transparency mandates on political parties, and decentralisation for diverse ideas.

This is an example of the application of corporate governance, and the refining of corporate laws to correct mismanagement and fraud. Considering that, several cases in the past like Sahara OFCD where corporate laws were manipulated to act on fraudulent activity. This highlights the crucial role of corporate governance in India, for the purpose of maintaining transparency, accountability, and ethical business practices.

Works Cited

- “Corporate Governance.” ICAEW,
<https://www.icaew.com/technical/corporate-governance/principles/principles-articles/does-corporate-governance-matter#:~:text=of%20the%20company.,Corporate%20governance%20is%20the%20system%20by%20which%20companies%20are%20directed,governance%20structure%20is%20in%20place>. Accessed 6 Aug. 2025.
- “How India Averted the Crisis and Liberalized Its Economy.” Google Arts & Culture,
<https://artsandculture.google.com/story/how-india-averted-crisis-and-liberalized-its-economy/2gURxpnXavp7Xg?hl=en>. Accessed 6 Aug. 2025.
- Comptroller and Auditor General of India. Union Government (Commercial) Report No. 8 of 2021 – Corporate Governance in CPSEs.
https://cag.gov.in/uploads/download_audit_report/2021/8_C-4%20Corporate%20Governance-061bb1b6d98c843.75300583.pdf. Accessed 6 Aug. 2025.
- Bhandari, Anandita. “The Haridas Mundhra LIC Scam and the Butterfly Effect.” *Mint*, 3 Jul. 2023,
<https://www.livemint.com/special-report/haridas-mundhra-lic-scam-financial-scamster-jawahar-lal-nehru-feroze-gandhi-bse-parliament-butterfly-effect-11726286671326.html>. Accessed 6 Aug. 2025.
- “Evolution of Corporate Governance in India.” LawBhoomi,
<https://lawbhoomi.com/evolution-of-corporate-governance-in-india/>. Accessed 6 Aug. 2025.
- Dalrymple, William. “Book Review: The Anarchy: The Relentless Rise of the East India Company.” *LSE Review of Books*, 14 Jan. 2020,
<https://blogs.lse.ac.uk/lsereviewofbooks/2020/01/14/book-review-the-anarchy-the-relentless-rise-of-the-east-india-company-by-william-dalrymple/#:~:text=In%20his%20latest%20work%20The,their%20rule%20over%20the%20subcontinent>. Accessed 6 Aug. 2025.
- Varma, Jayati. “Changing Perspectives of Corporate Governance in India.” *ResearchGate*, July 2014,
https://www.researchgate.net/publication/264441260_Changing_perspectives_of_corporate_governance_in_India. Accessed 6 Aug. 2025.

AI-Powered Early Warning System: Safeguarding Microenterprises Through Predictive Analytics By Vrinda Bisoi

Abstract

This study investigates the use of machine learning models to develop an early warning system for microfinance institutions (MFIs) to predict credit risk and ensure financial stability. Linear Regression, Random Forest, and XGBoost models were evaluated based on their performance in predicting loan defaults, using key indicators such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. The research found that Linear Regression underperformed across all metrics, showing weak predictive power. Both Random Forest and XGBoost performed well on training data but suffered from overfitting, leading to poor generalization on testing data. Feature importance analysis revealed that the number of loan officers (FA.MIX.LNOF), loan-to-officer ratio (FA.MIX.LPOF), and net impairment loss (FA.MIX.NILG) were critical factors influencing financial health and loan performance. The study concludes that while machine learning models hold potential for predicting credit risk in MFIs, further improvements are needed to enhance model generalizability and mitigate issues like overfitting and class imbalance.

Introduction

Microfinance institutions (MFIs) are organizations targeted towards financing small-scale enterprises and overall helping the economies of developing countries. The popularity of microfinance among socially responsible investors has meant that millions of extra poor have access to financial services today (Kannan and Panneerselvam). They differ from standard banks based on their usage of small loans in building micro credit. According to the Task Force on “Supportive Policy and Regulatory Framework for Microfinance”, microfinance is the “provision of thrift, credit and other financial services and products of very small amounts to the poor in rural, semi-urban or urban areas for enabling them to raise their income levels and improve living standards” (Pal et al.).

One of the largest markets for microfinance services is India, which has approximately 400 million people living in poverty. Because of this, there are nearly 75 million households that need microfinance (Fisher and Sriram). The perspective in India towards micro-finance has changed over time. Whereas earlier MF was “mandated credit”, today it covers a whole range of financial services required by the poor and is now viewed as a business proposition (Dasgupta and Rao). In 1987, National Bank for Agriculture and Rural Development (NABARD) provided Rs 10 lakh as grant assistance to Mysore Resettlement and Development Agency (MYRADA) to be used as seed money for its Credit Management Groups (CMGs). This successful experiment was the precursor to the launching of the pilot project in 1992 for linking 500 SHGs with banks. In February 1992, the project was made operational throughout the country. Thus, NABARD’s initiative brought in an additional constituent in the rural financial sector, i.e NGOs, voluntary organizations and NBFCs who have been promoting the self-help concept (Pal et al.).

The reliance on MFIs in developing countries such as India raises the question of whether the performance of these institutions can be trusted and its determinants for failure. Although microfinance is an important tool for economists to expand market for sustaining steady growth, it is a difficult proposition for formal financial institutions (FFI) due to high (i) travel and supervision costs to deal with large number of clients in infrastructurally backward areas, (ii) ledger cost to maintain large number of records and (iii) manpower cost to identify large number of clients, maintain their accounts and supervise their activities. The risk is high mainly because of the information gap. The information gap exists at two levels: (1) person to person, i.e. between borrower and lender and (2) person to environment, i.e. (a) between lender and the local socio-economic environment, and (b) borrower and production environment, i.e. technology, market, skill etc. (Dasgupta and Rao) Worldwide, approximately 2.5 billion people do not have a formal account at a financial institution, according to the World Bank's Global Financial Inclusion Database. As a result, most poor households operate almost entirely in the cash economy, particularly in the developing world (Pal et al.).

AI-powered early warning systems are methods being applied in business environments as a way to combat these challenges. One of the key applications of AI and ML in cloud security is anomaly detection. Machine learning algorithms can analyze patterns in cloud usage data, such as user behavior, network traffic, and resource utilization, to identify deviations that may indicate potential threats. By establishing a baseline of normal activity, AI-powered systems can detect and flag anomalies in real-time, enabling early detection and prevention of security incidents (Thamer). Advanced anomaly detection techniques, such as deep learning and unsupervised learning, can uncover complex patterns and identify subtle indicators of compromise that may be missed by traditional rule-based approaches. These AI-powered systems can continuously learn and adapt to evolving threat patterns, enhancing their detection capabilities over time (Thamer).

Predictive analytics can involve the use of techniques like supervised learning, time series analysis, and anomaly detection to forecast the likelihood of various security events, such as unauthorized access attempts, data breaches, and malware infections. By anticipating these threats, security teams can implement preventive measures, allocate resources more effectively, and make data-driven decisions to enhance the overall security posture of the cloud infrastructure (Thamer).

By observing the effects of predictive analytics in modern business environments correlated to the failure of micro-enterprises, we raise the main question of this study: "How can AI-powered early warning systems effectively safeguard micro-enterprises through predictive analytics?" The general objectives of this study are to build a predictive analytics model tailored for micro enterprises, assess the impact of AI-powered early warning systems on micro-enterprise sustainability, identify key risk indicators that predict micro-enterprise failure, and provide recommendations for implementing AI solutions in micro-enterprise risk management.

While there are many studies regarding micro-financing through a literary perspective, this study is unique as it develops a machine learning model that can be used to reduce failure

rates and enhance financial stability among micro-enterprises. The remainder of this paper is structured as follows: Section 2 provides a detailed literary review about existing literature and theoretical foundations regarding micro-enterprises and the key performance indicators that may contribute to their failure. Section 3 describes the data collection process as well as sources used from the MIX Data Market. Section 4 outlines the research methodology including the design of the machine learning model and analysis techniques. Section 5 provides the results of the study and a detailed discussion, addressing the broader implications of the results in the microfinance business sector. Finally, section 6 completes the paper by summarizing the research and suggesting ideas for future research in correlation with the study's results.

Literature Review

In recent decades, early warning systems and machine learning models have been commonly used as a tactic to enhance modern businesses. Several studies have provided a selection of common elements that should be included in the design and implementation of early warning systems (Segalini). These typically include the following components:

1. Field monitoring system, able to acquire physical quantities related to the phenomenon and transmit them;
2. Analysis and forecasting method, which describes the evolution, predicts its behavior, and identifies critical events on the basis of alert thresholds if available;
3. Warnings and dissemination of alert messages, to notify the relevant authorities about the forecasted critical event;
4. Emergency plans and measures, to be activated on the basis of the forecasting results obtained by applying the numerical model.

Regarding micro-enterprise development theories, a study found four indicators for economic factors such as employment, sales growth, income stability and return on investment. The four indicators for social factors include basic needs, social recognition, empowerment and freedom, and clusters for indicators for environmental factors such as water and energy use, waste and emission, waste management and hygiene factor. The findings also suggest a reflective scale of sustainability factor in the context of informal micro-enterprise (Khan and Quaddus).

Previous studies regarding the effects of micro-enterprises on economies display the positive effects it has long term for small businesses. A recent impact evaluation study (Asghari et al.) has compared the socio-economic conditions of 560 member households from 223 SHGs located in 11 states before and after (spanning over an average of 3-year period) their association with these groups. A positive impact has been observed on economic as well as social aspects of the members such as assets, income, access to credit, savings, self-confidence and behavioral patterns. The average value of assets per household, including livestock has registered an increase of 72.3 percent from the pre to post-SHG situation. About 58.6 percent of the sample households have registered an increase in assets (Khan and Quaddus).

We can identify the impact of machine learning models on micro-enterprises by evaluating existing case studies. For example, In the financial services sector, a major bank utilized machine learning models for fraud detection and prevention. By integrating supervised learning techniques such as decision trees and ensemble methods, the bank was able to analyze transaction patterns and identify anomalous behavior indicative of fraudulent activity. This approach led to a marked reduction in false positives and improved the detection rate of genuine fraud cases. The implementation of AI in this context underscores the critical role of high-quality training data and the need for continuous model updates to address evolving fraud tactics (Khan and Quaddus).

In a study done on Islamic micro-enterprises, an AI driven risk model was developed to determine whether an individual qualifies for an Islamic micro-financing loan. The model utilized a confusion matrix comparison, which displayed a strong correlation between monthly income and an individual's qualification for a loan. The model also evaluated an individual's criminality score, credit risk and used a random forest framework to estimate the maximum loan amount. The results of the study indicated that characteristics such as honesty and high charity contributions, measured by behavioral factors correlate with better likelihood of repayment. The implementation of an AI risk management model in this scenario provides a critical factor in strengthening governance and fairly evaluating micro-loan applicants (Katterbauer and Moschetta).

Although the provided studies expand on the benefits of utilizing AI and machine learning in micro-enterprises, numerous studies question the fairness of AI algorithms. While the recently published "European Code of Good Conduct for Micro-credit Provision" brings attention to ethical considerations when utilizing algorithms in the underwriting process, it is worth noting that the code's comments on algorithmic use in decision-making are somewhat limited. Although the code calls for providers to have a non-discrimination policy that specifies that discriminatory variables in algorithms should be excluded, even if they correlate with repayment likelihood, it may not go far enough in addressing the complexities of the notion of fairness (Koefer et al.).

Data

Data for this study came from the Microfinance Information Exchange (MIX) Market, a platform that collects financial and operational data from microfinance institutions (MFIs) and microenterprises. The variable names, as well as their descriptions are reported in Table 1. Initially, the data set included information from micro-enterprises in both Bangladesh and India. However, because India's data dominated the sample, we decided to focus exclusively on Indian data to avoid imbalance and ensure consistency in our analysis.

The data set included variables such as loan portfolio size, operational expenses, repayment rates, and loan loss rates, all of which were important to understanding microenterprise performance. During the data cleaning process, we dropped any rows that had no data at all to ensure that the analysis was based on complete records. For the response variable,

loan loss rate, we removed entries with invalid values (below 0% or above 100%) since these were errors.

Table 1: Financial Metrics Table

Code	Description
FA.MIX.LNLR	Loan loss rate
FA.MIX.LNSM	Loans per staff member
FA.MIX.LNOF	Number of loan officers
FA.MIX.PFEM.LC	Percent of female loan officers (%)
FA.MIX.PEXP.LP	Personnel expense / loan portfolio (%)
FA.MIX.NILG	Net impairment loss, gross loan portfolio
FA.MIX.LPOF	Loans per loan officer
FA.MIX.LNBB	Average loan balance per borrower

To handle missing values in the other variables, we used the median to fill in the gaps. This method was chosen because it is less affected by outliers and ensures that the distribution of the data is not significantly skewed. All variables were standardized to have a mean of 0 and a standard deviation of 1. This step was essential to ensure that variables on different scales did not influence the models.

In Figure 1, we can see that the correlation matrix is used to explore the relationships between numeric characteristics in the data set. This helped us identify which variables were strongly related to one another, with darker red indicating a strong positive correlation (close to 1) and darker blue showing a strong negative correlation (close to -1). Some features that have a strong correlation include FA.MIX.FIRE.LN and FA.MIX.GLOP, FA.MIX.LOAN.FE and FA.MIX.LOAN.MA, and FA.MIX.INLN.LN and FA.MIX.FIRE.LN. These correlations were important for feature selection because they allowed us to avoid including variables that could cause redundancy in our models.

Methodology

The data set was divided into 80% training data and 20% testing data to support the development and evaluation of effective models. This approach ensured that there was enough data available to train the models while keeping a separate portion of the data for testing. The training set was used to build and optimize the models, while the testing set provided an independent check on how well the models performed on new data.

A “traffic light system” was developed to categorically model loan loss rates, dividing them into three distinct categories: Green (0 - low loan loss rates), Yellow (1 - moderate loan loss rates), and Red (2 - high loan loss rates).

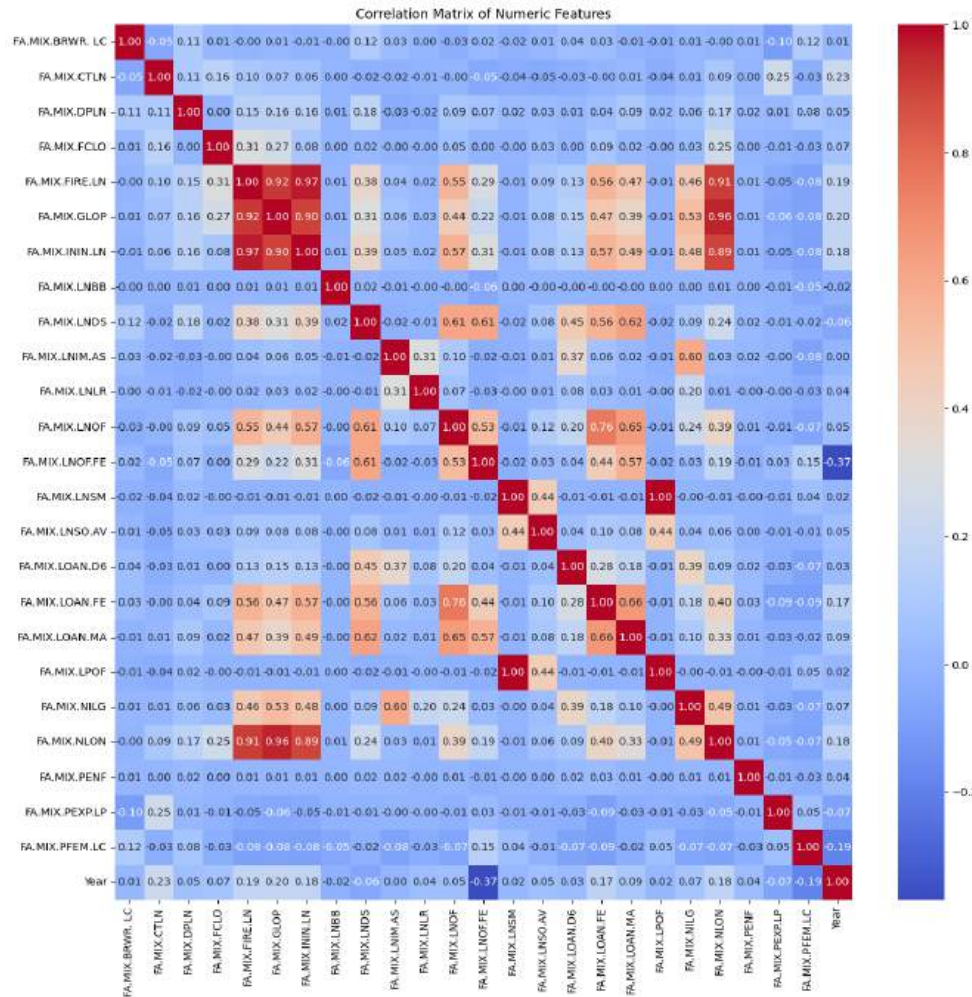


Figure 1: Correlation Matrix for Numeric Variables.

To model these categories, three classification algorithms were employed: Logistic Regression, a simple model that handles two data factors, Random Forest Classifier, a tree-based model that displays nonlinear relationships among features, and XGBoost Classifier, gradient-boosted decision trees that can handle complex data structures. The Synthetic Minority Oversampling Technique (SMOTE) was used to generate synthetic examples for the underrepresented Yellow and Red categories, creating a more balanced data set and reducing bias in model training. For continuous modeling of the loan loss rate, we applied three regression models: Linear Regression, Random Forest Regressor, and XGBoost Regressor

Linear regression served as the baseline for the model, providing a simple linear relationship between predictors and the target variable. Random Forest, a tree-based ensemble method, was used to identify nonlinear relationships in the data. XGBoost was used to further improve the predictive accuracy. To evaluate the performance of the model, the following metrics were used:

1. Mean Absolute Error (MAE): Measures how far the predictions are from the actual values on average, making it easy to understand the overall prediction accuracy.
2. Root Mean Square Error (RMSE): Shows the size of the prediction errors, giving more weight to larger errors, which makes it useful for identifying significant inaccuracies.
3. R-square (R^2): Indicates how well the model explains the variation in the actual data, showing how much of the changes in the response variable can be predicted by the model.

Findings

Treating the response as a categorical variable

The Multinomial Logistic Regression model achieved moderate precision, with 50% in the training set and 65% in the test set. The precision was highest for the “red” class at 0.03, reflecting the model’s challenge with the dataset’s class imbalance and the scarcity of “red” instances. Recall was strongest for the “green” class at 71%, showing the model could correctly identify most “green” instances, while the “yellow” and “red” classes had significantly lower recall. These results highlight the model’s bias toward the majority “green” class and its struggles with minority classes.

The Random Forest Classifier had 100% accuracy, but its testing accuracy dropped to 79%, indicating overfitting. Precision was highest for the “green” class at 92%, but both the “yellow” and “red” classes had poor precision. Recall followed a similar trend, with the “green” class achieving 84% recall, the “yellow” class 36%, and the “red” class 0%, reflecting the model’s inability to identify any “red” instances on the testing set. These results suggest that while the model performed well for the majority class, it struggled with minority class predictions.

The XGBoost Classifier had similar performance to Random Forest, with 100% accuracy on the training set and slightly higher testing accuracy at 80%. Precision was strong for the “green” class at 91%, while the “yellow” and “red” classes had much lower precision values of 23% and 0%, respectively. Recall for the “green” class was the highest at 86%, followed by the “yellow” class at 31%, and the “red” class at 0%. Despite slightly better testing accuracy, the XGBoost model faced similar challenges with class imbalance, particularly for the minority “red” class.

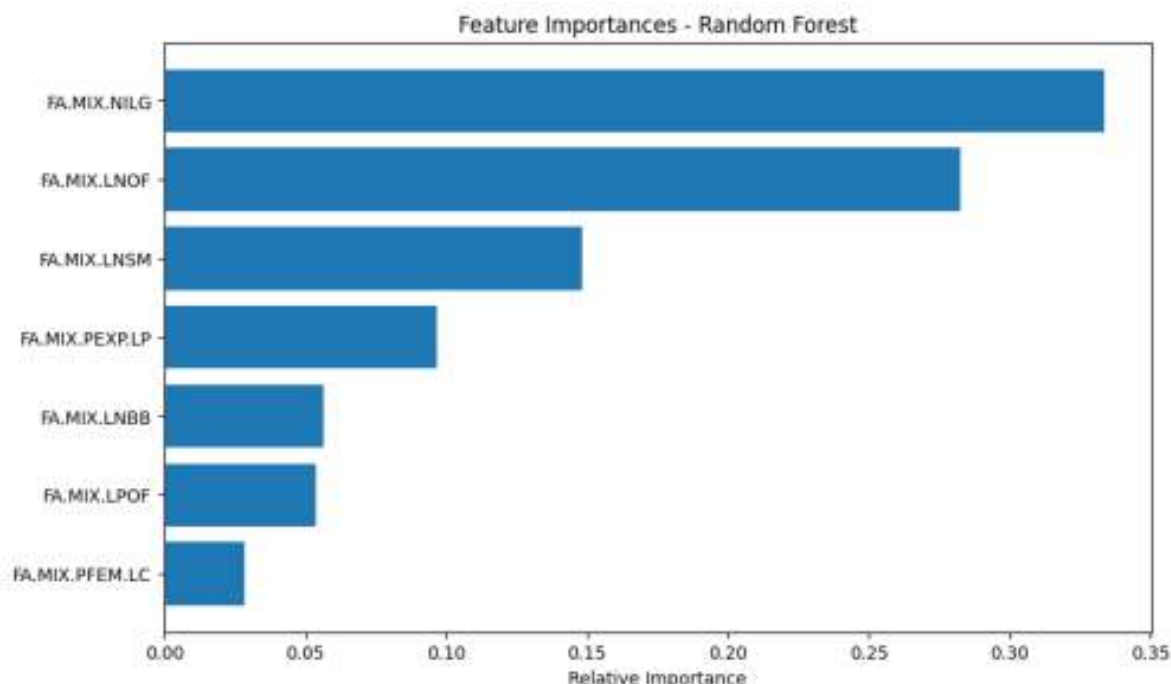


Figure 2: Feature Importance for Random Forest Classifier

The feature importance analysis for the Random Forest model under a categorical approach, as shown in (Figure 2) demonstrates that the three most influential features are FA.MIX.NILG (Net impairment loss, gross loan portfolio), FA.MIX.LNOF (Number of loan officers), and FA.MIX.LNSM (Loans per staff member).

The net impairment loss to gross loan Portfolio is the most influential predictor. This feature captures the proportion of impaired loans relative to the total loan portfolio, making it a critical indicator of financial risk. A higher value suggests that a significant portion of the loan is not being repaid, which signals financial instability. Its importance underscores the necessity of closely monitoring impairment losses to detect early warning signs and mitigate potential risks effectively.

The second most important feature is the number of loan officers, which emphasizes the human resource factor in risk management. Loan officers are directly involved in client management, repayment follow-ups, and assessing borrower risks. Enterprises with more loan officers likely benefit from better oversight, stronger relationships with borrowers, and a lower likelihood of default. Conversely, a limited number of officers might indicate excessive workloads, potentially weakening overall efficiency and increasing the chances of financial distress.

The loans per staff member ranks third in importance, reflecting the operational efficiency and workload distribution within the enterprise. A high ratio of loans to staff members suggests overburdened employees, which can lead to reduced engagement quality and weaker risk assessment. Balancing this metric is essential for maintaining both staff productivity and effective risk management. Together, these three features provide a well-rounded perspective on

financial and operational factors that significantly impact microenterprise performance, making them essential inputs for an AI-powered early warning system.

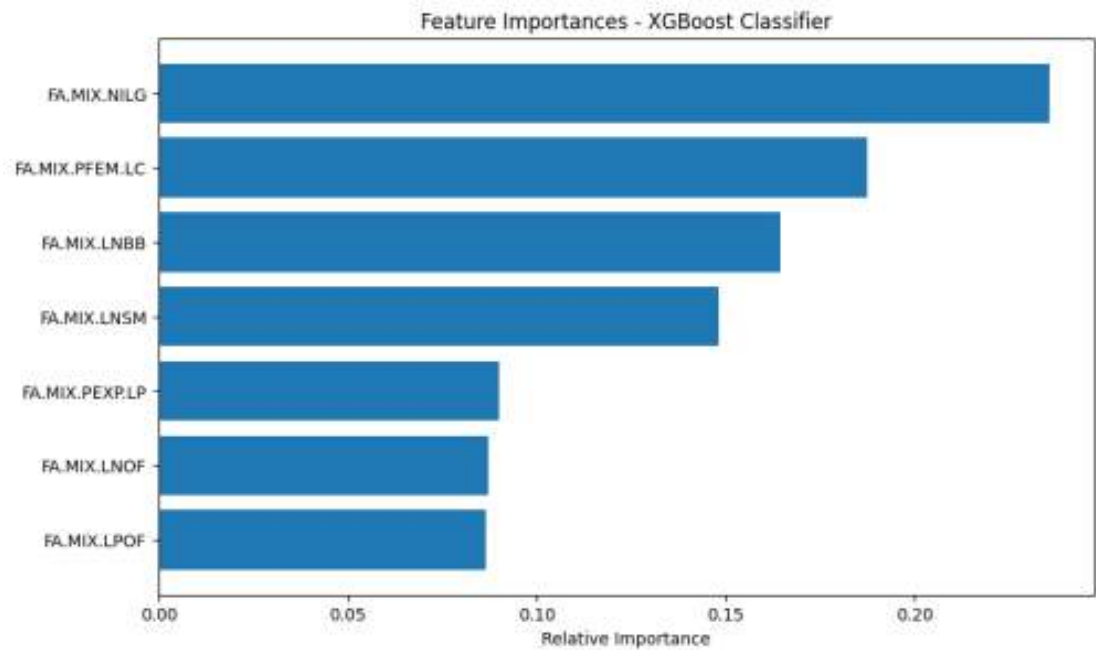


Figure 3: Feature Importance for XGBoost Classifier

The feature importance analysis for the XGBoost model under a categorical approach as shown in (Figure 3) demonstrates that the three most influential features were FA.MIX.NILG (Net Impairment Loss, Gross Loan Portfolio), FA.MIX.PFEM.LC (Percent of Female Loan Officers), and FA.MIX.LNBB (Average Loan Balance Per Borrower).

FA.MIX.NILG represents the financial health of lending institutions. Net impairment loss measures the portion of the loan portfolio that is expected to underperform or default, making it connected to an institution’s profitability and stability. High impairment losses suggest that the institution is dealing with a greater number of non-performing loans, indicating higher financial risk. This metric helps models assess credit risk effectively, as it captures key patterns associated with loan performance and repayment difficulties, which are essential for predicting financial instability or borrower defaults.

The second most important feature is the percent of female loan officers, which provides valuable insight into workforce diversity and its impact on lending practices. Gender diversity among loan officers may influence borrower behavior, trust, and overall customer experience. In some cases, a higher percentage of female officers can lead to stronger borrower relations or improved repayment rates, particularly in communities where borrowers feel more supported by a diverse staff. By incorporating this feature, models can better capture the subtle effects of diversity on loan performance and borrower behavior, which are often overlooked but highly relevant for predictive accuracy.

The third most important feature is FA.MIX.LNBB, which provides insight into the size and type of loans issued by an institution. Larger average loan balances often indicate a focus on

high-value clients or significant financial transactions, which typically involve greater risks. If a borrower defaults on a larger loan, the financial impact on the institution can be substantial, making this a key factor for evaluating risk. Loan size also tends to influence borrower behavior, with repayment patterns or default rates often varying based on the loan amount. By considering this feature, the model can better account for these differences, leading to more accurate assessments of creditworthiness and risk.

Treating the response as a continuous variable

In Table 2, Linear Regression, Random Forest, and XGBoost were evaluated on their performance using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score. These metrics provide insights into the models’ predictive accuracy and their ability to generalize to unseen data. The analysis presents both the training and testing set results for each model, allowing for an overview of their strengths and weaknesses.

Table 2: Model Performance Metrics for Linear Regression, Random Forest, and XGBoost

Model	Metric	Training Set	Testing Set
Linear Regression	MAE	0.94	0.79
	RMSE	4.13	1.49
	R^2 Score	0.04	-0.04
Random Forest	MAE	0.34	0.70
	RMSE	1.56	1.99
	R^2 Score	0.86	-0.85
XGBoost	MAE	0.07	0.72
	RMSE	0.11	2.51
	R^2 Score	1.00	-1.95

For the Linear Regression model, both training and testing set results indicating limited predictive power. On the training set, the model showed an MAE of 0.94 and an RMSE of 4.13, showing significant deviations from the actual values. The R^2 score of 0.04 indicated that the model explained only 4% of the variance in the data. The testing set showed a slight improvement in MAE (0.79) and a decrease in RMSE (1.49), but the model still performed poorly, as evidenced by the negative R^2 score of -0.04. This suggests that Linear Regression struggles to capture the underlying patterns in the data and fails to generalize well to unseen data.

Random Forest exhibited a strong fit to the training data but struggled with generalization to the testing set. On the training set, it achieved a low MAE of 0.34 and a high R^2 score of 0.86, indicating a good ability to capture the variance in the data. However, when applied to the testing set, the MAE increased to 0.70, the RMSE increased to 1.99, and the R^2 score fell to -0.85. These

results suggest that Random Forest overfitted the training data, memorizing its patterns and failing to adapt well to new data. The inability of the model to generalize highlights the need for techniques such as cross-validation or regularization to improve its performance on unseen data.

XGBoost achieved an MAE of 0.07 and a perfect R^2 score of 1.00, indicating a near-perfect fit to the training data. However, like Random Forest, XGBoost struggled with overfitting, as evidenced by its sharp decline in performance on the testing set. The MAE increased to 0.72, and the RMSE increased to 2.51, signaling significant errors in its predictions on unseen data. The negative R^2 score of -1.95, highlights the overfitting issue. Despite its strong training performance, XGBoost’s failure to generalize to new data suggests that further adjustments, such as hyperparameter tuning, are needed to improve its validity.

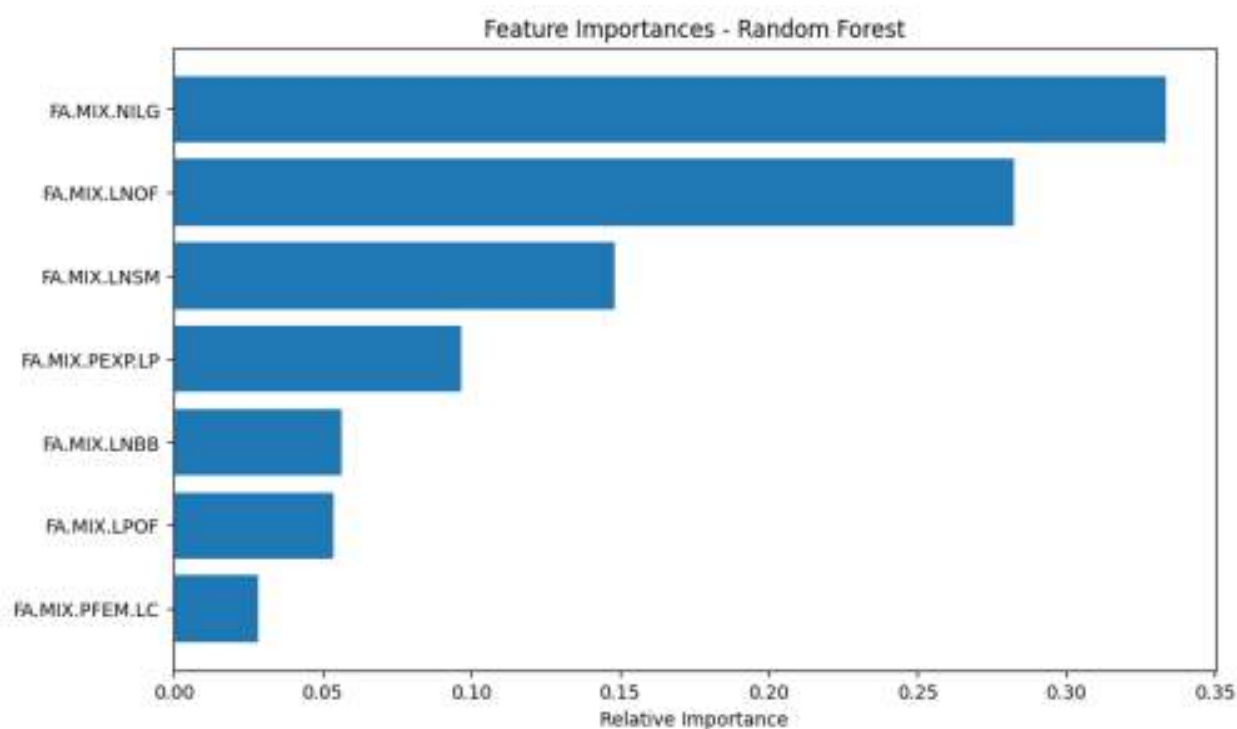


Figure 4: Feature Importance Random Forest

The three features that had the most significant impact in influencing the Random Forest model, shown in (Figure 4), are FA.MIX.NILG (Net impairment loss, gross loan portfolio), FA.MIX.LNOF (Number of loan officers), and FA.MIX.LNSM (Loans per staff member). FA.MIX.NILG, the most important feature, is the ratio of net impairment loss to the gross loan portfolio. This metric reflects the proportion of impaired loans relative to the total loan portfolio, making it a key indicator of financial risk. A higher value suggests that a significant portion of loans are not being repaid, which could signal underlying financial instability. This finding underscores the importance of closely monitoring impairment losses to identify early signs of distress.

The second most important feature was FA.MIX.LNOF, which is the number of loan

officers. Loan officers are essential for managing client relationships, ensuring timely repayment, and assessing risk. A higher number of loan officers can indicate better loan oversight and improved client management, reducing the likelihood of default. In contrast, insufficient staffing could lead to overburdened officers, which can weaken risk management processes. This result emphasizes the critical role that human resources play in the success and stability of micro-enterprises.

Finally, FA.MIX.LNSM represents the loans per staff member, ranked as the third most influential feature. This metric measures the workload distributed between staff, serving as a proxy for operational efficiency. A higher ratio might indicate that employees are handling too many loans, which can reduce the quality of client engagement and increase the likelihood of errors in risk assessment. Striking the right balance in this metric is vital to ensure both efficiency and sustainability. Together, these features highlight the importance of combining financial indicators with operational dynamics in an AI-powered early warning system tailored for micro-enterprises.

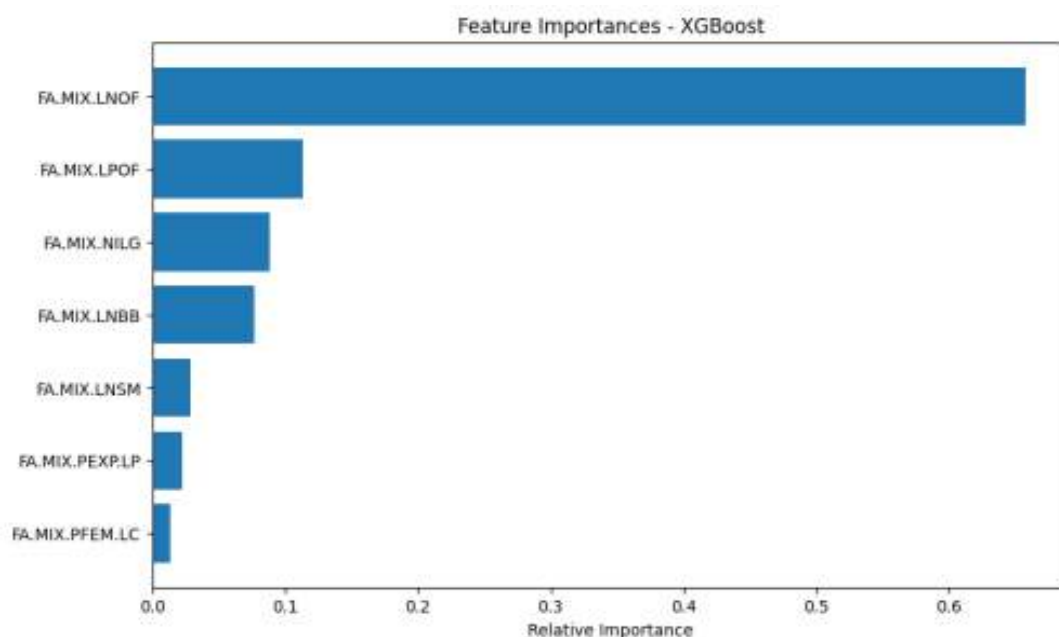


Figure 5: feature importance XGBoost

The feature importance analysis for the XGBoost (Figure 5) reveals that the top three features influencing the model's predictions are FA.MIX.LNOF (Number of loan officers), FA.MIX.LPOF (Loans per loan officer), and FA.MIX.NILG (Net impairment loss, gross loan portfolio). FA.MIX.LNOF, which is the most important feature, shows how crucial loan officers are in financial institutions. The greater the number of loan officers, typically indicates that the institution can handle more customers, credit evaluations, and loan management and directly helps the institution grow its loan portfolio while keeping credit risks under control.

FA.MIX.LPOF, the second most important feature, looks at the ratio of loans to loan officers. This gives insight into how efficiently loan officers are working. While a higher ratio

might show that the system is streamlined and productive, it could also indicate that loan officers are over loaded. If this happens, loan evaluations might suffer, leading to increased credit risk. This feature is a good measure of how well resources are being used and risks are being managed.

The third feature, FA.MIX.NILG, shows the net impairment loss compared to the gross loan portfolio. This is a key indicator of the quality of the loan portfolio. A higher value means there are more non-performing loans, which hurts profitability and financial stability. Because of this, it's an essential metric for understanding credit risk and the overall financial health of the institution.

Table 3: Performance Comparison of Multinomial Logistic Regression, Random Forests, and XG Boost Classifiers

Metric	Dataset	Classifier	Green	Red	Yellow	Overall
Precision	Training	Multinomial LR	0.42	0.58	0.60	0.53
		Random Forest	1.00	1.00	1.00	1.00
		XGBoost	1.00	1.00	1.00	1.00
	Testing	Multinomial LR	0.91	0.03	0.19	0.37
		Random Forest	0.92	0.00	0.23	0.38
		XGBoost	0.91	0.00	0.23	0.38
Recall	Training	Multinomial LR	0.68	0.63	0.18	0.50
		Random Forest	1.00	1.00	1.00	1.00
		XGBoost	1.00	1.00	1.00	1.00
	Testing	Multinomial LR	0.71	1.00	0.15	0.62
		Random Forest	0.84	0.00	0.36	0.40
		XGBoost	0.86	0.00	0.31	0.39
F1-Score	Training	Multinomial LR	0.52	0.61	0.28	0.47
		Random Forest	1.00	1.00	1.00	1.00
		XGBoost	1.00	1.00	1.00	1.00
	Testing	Multinomial LR	0.79	0.05	0.17	0.34
		Random Forest	0.88	0.00	0.28	0.39
		XGBoost	0.89	0.00	0.26	0.38
Accuracy	Training	Multinomial LR	0.50 (3759)			
		Random Forest	1.00 (3759)			
		XGBoost	1.00 (3759)			
	Testing	Multinomial LR	0.65 (355)			
		Random Forest	0.79 (355)			
		XGBoost	0.80 (355)			

Conclusions and Remarks

This study evaluated the performance of Linear Regression, Random Forest, and XGBoost in building an early warning predictive system for microfinance institutions (MFIs). The results showed that linear regression performed poorly, with low R^2 scores and high error rates in both the training and testing sets. Random Forest and XGBoost demonstrated strong predictive power in training data, but suffered overfitting, leading to poor generalization in testing data. The feature importance analyzes highlighted key factors that influence financial

health, including FA.MIX.LNOF (number of loan officers), FA.MIX.LPOF (loans per loan officer), and FA.MIX.NILG (net impairment loss, gross loan portfolio), underscoring their importance in predicting credit risk and institutional stability. While these models offer potential for risk prediction, the results suggest that further improvements are necessary to enhance their generalizability.

The main limitation of this study lies in the inability of the models to effectively generalize to unseen data, as shown by the decline in performance in the testing sets. This issue was particularly evident with Random Forest and XGBoost, which overfitted the training data. Furthermore, the class imbalance in the data set negatively affected the ability of the models to predict minority classes, particularly the 'red' category, which consistently showed low or zero recall. Addressing these challenges would require techniques like resampling, hyperparameter tuning, and cross-validation methods. Another limitation is the potential lack of external validity, as the dataset used may not fully represent MFIs in other regions or contexts. The findings provide actionable information for policy makers and MFIs. Metrics like the number of loan officers and loan-to-officer ratios highlight the importance of efficient resource allocation in managing credit risk and improving operational performance. FA.MIX.NILG, which reflects net impairment losses, serves as a critical indicator of financial stability and could help institutions monitor and mitigate risks associated with non-performing loans. Additionally, the impact of gender diversity among loan officers (FA.MIX.PFEM.LC) suggests that workforce diversity plays a role in borrower behavior and repayment rates, which could inform hiring practices and training programs. However, implementing AI-based predictive models should be approached cautiously, as issues like overfitting and bias could lead to inaccurate predictions and unintended consequences for borrowers and institutions.

Future research should focus on improving model generalizability by addressing overfitting and class imbalance, possibly through advanced techniques like regularization, ensemble methods, or cost-sensitive algorithms. Testing these models on diverse datasets from different regions would also help assess their external validity and applicability across various MFIs. Moreover, integrating qualitative data, such as borrower demographics or repayment histories, could improve model accuracy and provide a more comprehensive understanding of credit risk. Lastly, exploring the real-world feasibility of implementing these predictive systems, including cost-benefit analyses and ethical considerations, would ensure that these tools are impactful for MFIs.

Works Cited

- Segalini, Andrea, et al. "Innovative Monitoring Tools and Early Warning Systems for Risk Management: A Case Study." *Geosciences*, vol. 9, no. 2, 29 Jan. 2019, p. 62, <https://doi.org/10.3390/geosciences9020062>. Accessed 17 Oct. 2024.
- Dasgupta, Rajaram, and K Dinker Rao. *Microfinance in India: Issues, Challenges and Policy Options*. Vol. 27, no. 2, 1 Jan. 2003, www.researchgate.net/publication/241672642_Microfinance_in_India_Issues_Challenges_and_Policy_Options. Accessed 17 Oct. 24AD.
- "Beyond Micro-Credit: Putting Development Back into Micro-Finance - Oxfam Policy & Practice." Oxfam Policy & Practice, 2023, policy-practice.oxfam.org/resources/beyond-micro-credit-putting-development-back-into-micro-finance-122837/. Accessed 22 Oct. 2024.
- Kannan, M., and A. Panneerselvam. "The Microfinance in India- an Overview." *International Journal of Current Research and Academic Review*, vol. 1, no. 2347-3215, Jan. 2013, scispace.com/papers/the-microfinance-in-india-an-overview-3piiyjwns8. Accessed 3 Nov. 2024.
- Klemens Katterbauer, and Philippe Moschetta. "A Deep Learning Approach to Risk Management Modeling for Islamic Microfinance." *DOAJ (DOAJ: Directory of Open Access Journals)*, 1 July 2022, ojs.unito.it/index.php/EJIF/article/view/6202, <https://doi.org/10.13135/2421-2172/6202>. Accessed 3 Nov. 2024.
- Khan, Eijaz Ahmed, and Mohammed Quaddus. "Development and Validation of a Scale for Measuring Sustainability Factors of Informal Microenterprises – a Qualitative and Quantitative Approach." *Entrepreneurship Research Journal*, vol. 5, no. 4, 1 Jan. 2015, <https://doi.org/10.1515/erj-2014-0017>. Accessed 7 Nov. 2024.
- Franziska Koefer, et al. "Fairness in Algorithmic Decision Systems: A Microfinance Perspective." *University of Twente Research Information*, Apr. 2023, research.utwente.nl/en/publications/fairness-in-algorithmic-decision-systems-a-microfinance-perspecti. Accessed 12 Nov. 2024.
- Pal, Suresh, et al. *Institutional Change in Indian Agriculture*. No. 11869, 1 Jan. 2003, www.researchgate.net/publication/23513793_Institutional_Change_In_Indian_Agriculture. Accessed 21 Nov. 2024.
- Asghari, Farhad, et al. "Comparative Study of the Financial Independence, Self-Confidence and Decision-Making Power, Awareness of the Social Resources and Mental Health in the Women Members of SHG and Non-Member of SHG." *Sociology Mind*, vol. 03, no. 01, 2013, pp. 114–117, <https://doi.org/10.4236/sm.2013.31017>. Accessed 26 Nov. 2024.
- Thamer Abdel-Wahid, and Research Scholar II. "AI-POWERED CLOUD SECURITY: A STUDY on the INTEGRATION of ARTIFICIAL INTELLIGENCE and MACHINE LEARNING FOR..." *ResearchGate*, unknown, 20 May 2024, www.researchgate.net/publication/383095008_AI-POWERED_CLOUD_SECURITY_A_STUDY_ON_THE_INTEGRATION_OF_ARTIFICIAL_INTELLIGENCE_AND_MACHINE_LEARNING_FOR_IMPROVED_THREAT_DETECTION_AND_PREVENTION. Accessed 10 Dec. 2024.

Rate, Load, and Medium Effects on Hysteresis in Coiled NiTi Springs

By Agastyaa Poddar

Abstract

Nickel–Titanium (NiTi) shows reversible thermo-elastic transformations that produce shape recovery and super-elasticity. This paper documents heating–cooling hysteresis in coiled NiTi springs using a low-cost rig suitable for a school laboratory. One end of the spring is clamped; a water-filled bottle applies an approximately constant bias mass through a smooth idler. Temperature near mid-span and axial length are recorded in the smartphone frame. The protocol varies heating rate (by hair-dryer distance), bias load (100 g and 200 g), and thermal medium (air vs. a brief 60 °C–70 °C mug hold), with three cycles per condition. From the video we construct length–temperature (L – T) loops, estimate apparent transformation temperatures using a percent-of-stroke rule, and compute a unit-less hysteresis index that normalizes loop area by temperature span and stroke. Across runs, faster heating widens the loop and slightly elevates apparent A_f ; higher load shifts A_s upward; repeated cycling narrows the loop; and a mug hold reduces loop width.

Keywords Nickel-Titanium, super-elasticity, Temperature, hysteresis index

1. Introduction

NiTi undergoes a reversible thermoelastic transformation between B2 austenite and B19' martensite, with an R-phase in some alloys (Otsuka and Wayman). When a coiled spring made from NiTi is held under a bias mass and heated, its length follows one path on heating and a different path on cooling, forming a hysteresis loop in the L – T plane. As Duerig, Melton, et al. note, spring geometry trades off force and displacement in ways that suit low-power actuation.

This study asks four practical questions suited to a teaching lab:

- How does heating rate affect loop shape?
- How does a change in bias mass shift the loop?
- Does the loop stabilize over repeated cycles?
- Does a brief near-isothermal hold reduce thermal lag compared to air heating?

The aim is descriptive: provide a transparent protocol, clean plots, and a small set of metrics that generalize across simple rigs.

2. Related Work

Transformation temperatures (M_s , M_f , A_s , A_f) depend on chemistry, heat treatment, and prior cycling. Rate effects and thermal gradients can widen loops and shift apparent temperatures. Under uniaxial load, transformation stresses vary with temperature approximately linearly (often discussed via a Clausius–Clapeyron-type slope). Coiled geometries magnify

stroke and are common in practical actuators, yet most introductory demonstrations use straight wires. Early cycles often show stabilization: loop area drops slightly as the microstructure settles. NiTi's reversible transformation and actuator use cases are well documented in standard references.(Otsuka and Wayman; Duerig, Melton, et al.; Lagoudas).

3. Materials and Methods

3.1 Materials

NiTi wire of diameter 0.50 mm was coiled to an outer diameter near 10 mm; test segments of coiled length ~ 22 cm were used. A string over a smooth pen/spoon routed the load to a water bottle (bias masses 100 g and 200 g). A kitchen thermometer probe recorded temperature near mid-span and a ruler provided length.

3.2 Apparatus and framing

One end of the spring was clamped between padded books using binder clips. The probe sat 1 cm–2 cm from the coils, not touching metal. The camera framed the spring, ruler, and thermometer display.

3.3 Procedure & Design of experiments

Factors: (i) rate via hair-dryer distance: *near* (5 cm–6 cm), *med* (12 cm–15 cm), *far* (25 cm–30 cm); (ii) bias mass: 100 g, 200 g; (iii) medium: air (all rates) and mug-hold (60 °C–70 °C) at 100 g, med. Three cycles per condition.

Run set: 100 g $\times$ (near, med, far) $\times$ cycles 1–3; 200 g $\times$ med $\times$ cycles 1–3; 100 g $\times$ med $\times$ mug $\times$ cycles 1–3.

Each run began with 60 s at room temperature to establish L_0 . Heating proceeded to ~ 75 °C–80 °C (or a 60 °C–70 °C hold in the mug condition), followed by passive cooling. Temperature and length were read every 2–3 s from the video.

3.4 Data and metrics

Recorded Metrics

- `time_s` — a common clock so temperature and length can be paired point-by-point.
- `temp_C` — the driver of the phase change; we use the probe *near* the spring so it responds fast enough to show heating vs. cooling differences.
- `length_mm` — the output we care about; this is what moves in an actuator. We read clamp $\rightarrow$ knot distance on a ruler in the video frame.
- `load_g` — the bias mass. More mass means more mechanical bias on the spring; it tends to shift the transformation temperatures and can change the stroke.
- `rate_level` and `medium` — how quickly/how uniformly we heat (air near/med/far vs. brief mug hold). These affect thermal lag and loop width.

- cycle — early cycles often “settle” the material; we keep cycles 1–3 separate to see that stabilization.
- IDs and geometry (run_id, spring_id, segment_len_cm, wire_d_mm, coil_OD_mm) — so results are traceable and comparable across pieces.

We turn the raw table into a heating curve and a cooling curve:

1. Sort each run by time_s. Find the time of peak temperature; everything before it is *heating*, everything after is *cooling*.
2. Smooth the length_mm series with a short moving average (5–9 points) to reduce ruler noise without hiding real changes.
3. Because heating and cooling may cover slightly different temperature ranges, we compare them only on their *shared* span $[T_{min}, T_{max}]$. We linearly interpolate both branches onto the same set of temperatures in that span. This lets us subtract one curve from the other fairly.
4. Plot **Length (mm)** vs **Temperature (°C)** for heating (orange) and cooling (blue). The loop between them is the hysteresis.

Hysteresis index

We want a single number that grows when the loop is “wider,” but does not depend on our units or on exactly how hot we went. We use a normalized area:

$$H = \frac{1}{(T_{max} - T_{min}) \Delta L_{max}} \int_{T_{min}}^{T_{max}} |L_{\uparrow}(T) - L_{\downarrow}(T)| dT,$$

where $L_{\uparrow}$ is heating, $L_{\downarrow}$ is cooling, and ΔL_{max} is the total length change (stroke) across that same temperature span.

Why does this help? The numerator is the shaded area between the two curves (bigger area = stronger hysteresis). Dividing by the temperature span and by the stroke makes the number scale-free, so two students who heat to slightly different peaks can still compare results. Typical values are roughly 0.3–0.5.

Apparent transformation temperatures (A_s, A_f, M_s, M_f)

Exact picks from noisy classroom data can be tricky if we look for tiny slope breaks. Instead we use a simple “percent-of-stroke” rule that is stable and easy to repeat:

1. On the *heating* branch, take the early length L_{start} and late length L_{end} (e.g., medians over the first and last few points). Define a progress fraction

$$f_h(T) = \frac{L(T) - L_{start}}{L_{end} - L_{start}}.$$

2. If the spring *contracts* on heating, flip the fraction so it still runs from 0 to 1.

3. Define A_s at 10% progress and A_f at 90% progress by reading off the temperatures where $f_h(T) = 0.10$ and 0.90 .
4. Repeat on the *cooling* branch to get M_s (10%) and M_f (90%).

Duerig, Pelton, et al. discuss why A_f alone is insufficient to predict plateau stresses across test protocols.

Why 10%/90%? It avoids edge noise at 0%/100%, it does not require derivatives, and it tracks the “useful” part of the change the device would feel. Report these as °C to one decimal place.

Experimental Limitations

- Thermometer resolution (e.g., ± 0.3 °C), ruler reading (± 0.5 mm), and video sampling (every 2–3 s).
- For each condition, report mean $\pm$ standard deviation across cycles (e.g., $A_s = 75.8 \pm 0.3$ °C).
- If two conditions differ by less than the combined uncertainty, say so plainly.

3.5 Clausius–Clapeyron relation

In NiTi, adding mechanical bias generally shifts the transformation to higher temperatures. We test that by checking whether the apparent austenite-start temperature A_s rises when we increase the hanging mass.

Following Lagoudas, the Clausius–Clapeyron slope can be written as $d\sigma/dT = \Delta s/\Delta \epsilon^{tr}$.

For a thermoelastic martensitic transformation under (straight-wire) uniaxial stress σ ,

$$\frac{d\sigma}{dT} = \frac{\Delta s}{\Delta \epsilon^{tr}} = \frac{\Delta h}{T \Delta \epsilon^{tr}},$$

where Δs is the entropy change per unit volume, Δh is the enthalpy change per unit volume, T is absolute temperature, and $\Delta \epsilon^{tr}$ is the transformation strain. In practice, start/finish stresses are often close to linear with temperature,

$$\sigma_{As}(T) \approx \sigma_0 + C(T - T_0), \quad \sigma_{Ms}(T) \approx \sigma'_0 + C'(T - T_0),$$

with slopes $C = d\sigma/dT$ on the order of a few MPa K^{−1} in NiTi. Inverting gives the temperature shift per stress, $dT/d\sigma = 1/C$. Building on the 1D constitutive framework of Brinson, we treat the heat-driven length change as a transformation strain proxy.

3.5.1 Applying it here (coiled spring, hanging mass).

Our rig does not measure σ in a straight wire. We hang a *mass* on a *coiled* spring, so axial force is turned into torsion and bending in the wire. The stress state depends on coil index, pitch, and active-coil count. Because of that geometry, we report an *apparent* C–C trend: when bias mass increases from 100 g to 200 g, A_s should shift upward. If wire diameter d and mean coil diameter D are known, an order-of-magnitude wire shear stress can be estimated by a standard spring relation,

$$\tau \approx \frac{8FD}{\pi d^3} \quad (\text{ignoring small correction factors}),$$

but non-uniform heating and curvature mean this is only approximate.

For medium-rate, air-heated runs (Fig. 3), we compute A_s per cycle using the 10%

heating-branch rule, average across cycles at each load, and then report:

1. $A_s(100 \text{ g})$ and $A_s(200 \text{ g})$ as mean $\pm$ SD;
2. the difference $\Delta A_s = A_s(200) - A_s(100)$ in $^{\circ}\text{C}$;
3. the slope in $^{\circ}\text{C}$ per 100 g, i.e. ΔA_s divided by 100 g.

If a load cell and geometry are later added, the same plot can be recast as A_s versus stress, and the slope converted to $dT/d\sigma$ for comparison with textbook C–C values.

We avoid converting grams to MPa in this paper. Temperature is measured *near* the spring, not inside the wire; the 10% rule yields *apparent* A_s values that depend on the analyzed temperature span and smoothing. These limits are stated alongside the results.

4. Results

A montage of representative L–T loops is shown in Fig. 1. We then quantify four factors: *rate*, *load*, *training*, and *medium*. For clarity, we first report cycle-averaged effect sizes; the subsections that follow provide the full overlays and details.

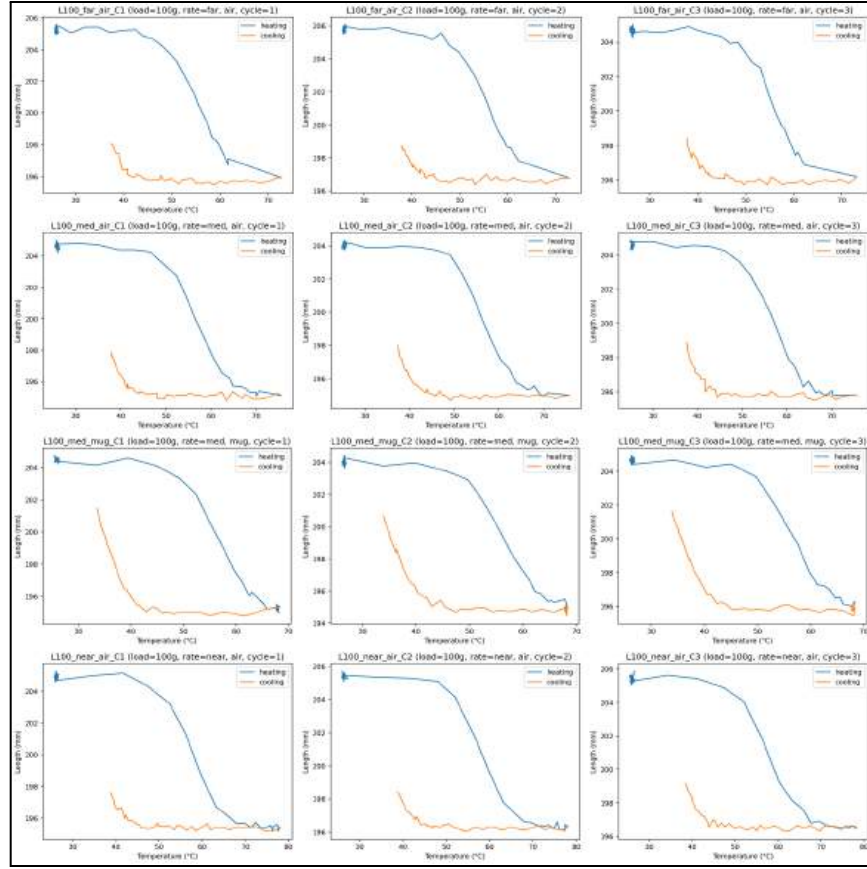


Fig 1: Representative length–temperature loops across conditions. Heating (blue) and cooling (orange) are plotted for each run.

Unless noted, values are means $\pm$ SD across the three cycles at each condition.

- **Rate (Fig. 2).** At 100 g, cycle 1, A_s was higher for *near* versus *far* heating (approximately 78.2°C vs. 72.8°C); $\Delta A_s \approx + 5^\circ\text{C}$.
- **Load (Fig. 3).** For medium-rate air heating, A_s increased from $75.74 \pm 0.06^\circ\text{C}$ at 100 g to $76.21 \pm 0.10^\circ\text{C}$ at 200 g; $\Delta A_s = + 0.47^\circ\text{C}$ (i.e., $\approx + 0.47^\circ\text{C}$ per 100 g).
- **Training / Stabilization (Fig. 4).** At 100 g, medium, air, the hysteresis index decreased from $H_1 = 0.402$ to $H_3 = 0.392$; $\Delta H_{1 \rightarrow 3} = - 0.010$.
- **Medium (Fig. 5).** At 100 g, medium, $H_{air} = 0.40 \pm 0.01$ versus $H_{mug} = 0.33 \pm 0.01$; $\Delta H = - 0.07$ (about $- 17\%$ relative).

4.1 Heating rate at 100 g (air)

Fig. 2 overlays near/med/far heating at 100 g. The near condition shows the largest separation between branches and a higher apparent A_f , consistent with thermal lag at faster transients.

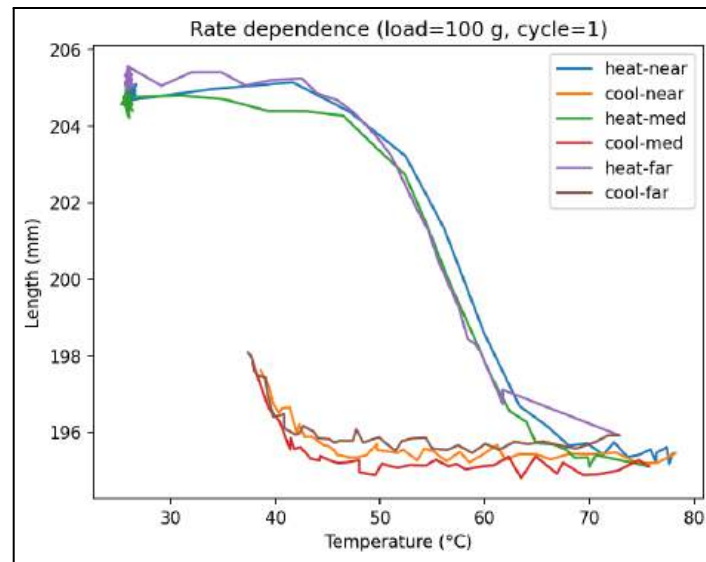


Fig 2: Rate dependence at 100 g, cycle 1. Hair-dryer distances: near (5 -6 cm), med (12 -15 cm), far (25 -30 cm).

4.2 Bias load at medium rate (air)

Fig. 3 plots A_s vs. load for medium-rate air-heated runs. The 200 g point lies above 100 g, indicating an apparent upward shift in A_s with increased bias mass, consistent with the C-C framing.

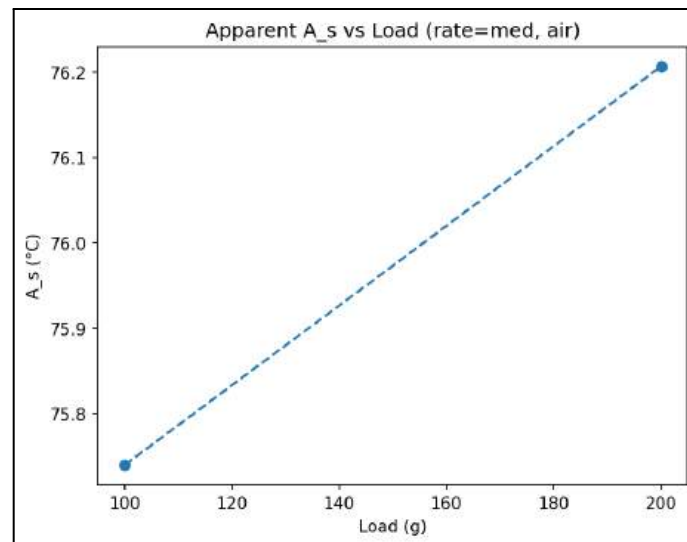


Fig 3: Apparent A_s vs. load for medium-rate, air-heated runs.

4.3 Stabilization across cycles

Fig. 4 shows the hysteresis index H from cycle 1 to 3 at 100 g, medium rate, air. H decreases slightly, consistent with early-cycle training. (An exponential fit may be added if desired.)

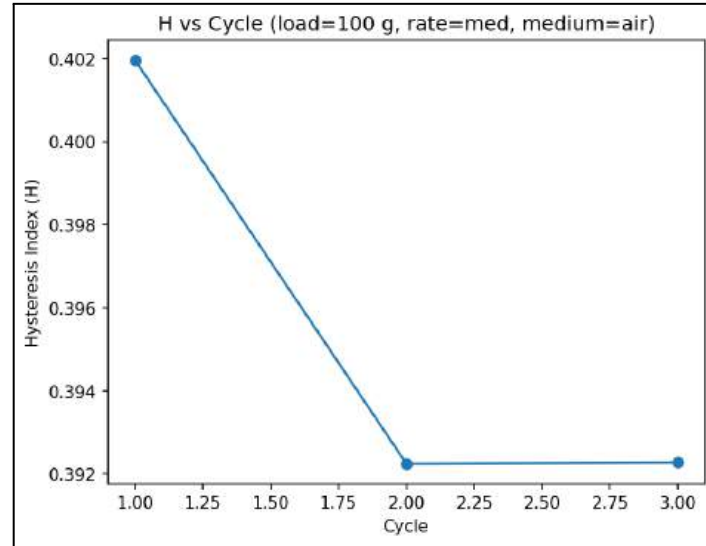


Fig 4: Hysteresis index H across cycles for 100 g, medium, air.

4.4 Medium: air vs. mug hold

Fig. 5 compares air heating with a brief mug hold at 60 --70 °C (both 100 g, medium rate, cycle 1). The mug loop is narrower and reaches its plateau earlier; a bar inset of H can summarize the reduction.

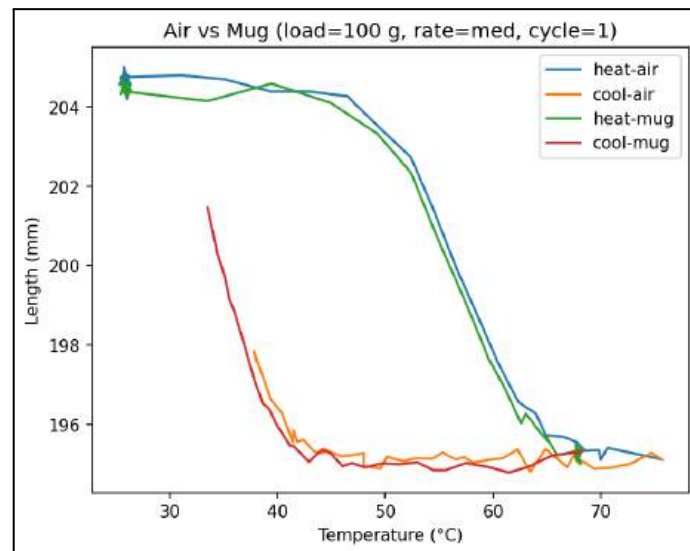


Fig 5: Air heating vs. mug hold at 100 g, medium rate, cycle 1. The mug condition reduces loop width (smaller H).

Table 1. Run metrics*Per-run metrics for NiTi spring hysteresis.*

load (g)	rate	medium	cycle	Tmin (°C)	Tmax (°C)	stroke (mm)	A_s (°C)	A_f (°C)	M_s (°C)	M_f (°C)	H
100	far	air	1	37.41	72.84	7.86	72.84	72.84	43.33	38.99	0.400
100	far	air	2	37.80	72.91	7.52	72.91	72.91	44.66	38.92	0.400
100	far	air	3	37.74	73.22	7.17	73.22	73.22	45.05	38.84	0.390
100	med	air	1	37.79	75.69	9.38	75.69	75.69	45.29	39.35	0.400
100	med	air	2	37.33	75.80	8.81	75.80	75.80	45.05	38.76	0.390
100	med	air	3	37.55	75.73	8.74	75.73	75.73	45.44	39.44	0.390
100	near	air	1	38.59	78.23	9.42	78.23	78.23	45.85	40.34	0.380
100	near	air	2	38.68	78.13	9.05	78.13	78.13	45.89	40.01	0.400
100	near	air	3	38.44	78.31	8.88	78.31	78.31	45.95	39.80	0.380
200	med	air	1	38.37	76.32	8.48	76.32	76.32	44.72	39.66	0.400
200	med	air	2	38.01	76.23	7.94	76.23	76.23	45.28	39.44	0.380
200	med	air	3	39.13	76.07	7.78	76.07	76.07	47.32	40.39	0.400
100	med	mug	1	33.50	68.30	9.26	68.30	68.30	41.71	35.10	0.320
100	med	mug	2	33.93	68.37	8.81	68.37	68.37	43.30	35.40	0.330
100	med	mug	3	33.89	68.45	8.38	68.45	68.45	42.48	35.54	0.330

Table 2. Design of Experiments (DoE)

load_g	rate_level	medium	n_cycles	cycles	n_runs
100	near	air	3	1,2,3	3
100	med	air	3	1,2,3	3
100	far	air	3	1,2,3	3
200	med	air	3	1,2,3	3
100	med	mug	3	1,2,3	3

5. Conclusion

The four comparisons answer the questions posed in the introduction. Faster heating widens loops and elevates apparent A_f ; increased bias mass raises A_s ; repeated cycles reduce H ; a brief near-isothermal hold narrows the loop. The load trend is consistent with the Clausius–Clapeyron framing in Eq. [eq:CC]: a higher mechanical bias requires a higher temperature to complete the austenitic transformation. Because coil geometry mixes torsion and bending in the wire, we report bias mass and avoid converting to MPa; this limitation is explicit.

One of the limitations is that stress is not measured directly; the thermometer reads near-wire temperature; ruler readings limit length precision; the percent-of-stroke rule yields *apparent* temperatures that depend on span and smoothing. Straightforward additions include a small load cell, a second probe on the cool side, and video tracking for sub-millimetre length. Geometry studies could vary active-coil count and pitch to compare stroke-per-coil and force–displacement trade-offs. A simple tabletop protocol yields reproducible L–T loops for NiTi springs and captures the effects of rate, load, cycling, and medium. The figures are sized for single-column print, and the dataset/scripts permit replication by other student groups.

Works Cited

- Brinson, L. Catherine. "One-Dimensional Constitutive Behavior of Shape Memory Alloys: Thermomechanical Derivation with Non-Constant Material Functions and Redefined Martensite Internal Variable." *Journal of Intelligent Material Systems and Structures*, vol. 4, no. 2, 1993, pp. 229–42, <https://doi.org/10.1177/1045389X9300400213>.
- Duerig, T. W., K. N. Melton, et al., editors. *Engineering Aspects of Shape Memory Alloys*. Butterworth-Heinemann, 1990.
- Duerig, T. W., A. R. Pelton, et al. "The Measurement and Interpretation of Transformation Temperatures in Nitinol." *Shape Memory and Superelasticity*, vol. 3, Nov. 2017, pp. 485–98, <https://doi.org/10.1007/s40830-017-0133-0>.
- Lagoudas, Dimitris C., editor. *Shape Memory Alloys: Modeling and Engineering Applications*. Springer, 2008, <https://doi.org/10.1007/978-0-387-47685-8>.
- Otsuka, Kazuhiro, and C. M. Wayman, editors. *Shape Memory Materials*. Cambridge University Press, 1998.