|CURIEUX|

ACADEMIC REVIEW

September Issue

Issue 11 Part 2

Editing Staff

Chief-In-Editor

George Hasnah

Lead Editor

Amanuel Geremew

Ashley Du

Assisting Editors

Jing Yan Huang

Nidhi Potineni

Sophia Kong

Tony Yuan

Ridit Aggarwal

Manu Sharma

Rishi Gadikota

Eric Wang

Table Of Contents

Page 6: The Legacy of Riot Grrrl: Musician Career Evolution and Contemporary Perspectives By Fern Preppernau

Page 20: The Paradox of Underrepresentation in Female Leadership By Yeryeong Chun

Page 40: A Comparative Modal Analysis of Automotive Suspension Systems By Yueqian Jiang

Page 56: Psychodynamic Psychotherapy In an Empirical World: A Literature Review of Clinical and Neuroscientific Efficacy By Gursher Chopra

Page 63: Nero and the Great Fire of Rome, 64 CE By Lily Li

Page 82: Melatonin Inhibits Human Leukemia U937 Cell Proliferation Via ERK Signaling Pathway Suppression By David Niu

Page 97: The Role of Financial Influencers and Social Media Streamers in Executing Pump-and-Dump Schemes Across Memecoin and Meme Stock Markets By Dev Kalaria

Page 107: The Effect of Demographics on Organic Food Perception By Venkata Anish Neti

Page 129: The Influence of Cognitive Versus Experiential Knowledge on Investing Behaviors Across Demographics By Mahika Shah

Page 151: The Effect of Global Financial Crisis and COVID-19 Pandemic on Return on Assets: Evidence From Stock Exchange of Thailand By Torfan Wantarnagon

Page 166: Understanding How Sneakers Act as a Social Currency Among Adolescents in India By Dhairya Dalani

Page 177: Global Stagflation and International Relations By Mikul Jain

Page 188: Comparing AI and Human Decision-Making in Home Loan Interest Rate Predictions By Charishma Konda

Page 195: Evaluating the Generalizability of an AI-Based Patient-Trial Matching Model in Glioblastoma By Aaron Vinoth

Page 210: A Study of Distance to Criticality in Conservative Zhang Sandpile Models of Random Neural Networks By Kean Li and Afshin Montakhab

Page 221: A Testable Framework for Platform-Driven Stock Volatility: Reddit Coordination vs. X Broadcast Effects Across Firm Capitalization By Aayan Sheriff

Page 234: The Effects of Western-Style and Eastern-Style Schooling and Parenting on Adolescent Mental Wellbeing: A Qualitative Study in Shanghai By Yu Pei Chen

Page 260: A Low-Cost, Multi-Functional System for Real-Time Equine Health Monitoring By Qiaoyi Chen

Page 270: Development and Biodegradability Assessment of Eco-Friendly Bioplastics Derived from Banana Peel Waste By Jiatong Xie

Page 279: The Effects of Acute Noise Exposure on Female Phonotaxis and Male Calling Behavior in *Gryllodes sigillatus* By Aiden Noh

Page 291: Dealing with an Asteroid Strike: The Impacts of the COVID-19 Pandemic and 2023 WGA and SAG-AFTRA Labour Stoppages on Toronto's Film Industry By Leila Popatia

Page 309: The Effects of Carbohydrate-Based and Ketogenic Diets on Energy Levels, Focus, and Sleep: A Randomized Single-Subject Crossover Study By Fernando Leyva-Duenas

Page 316: Impact of Ayurvedic Herbal Extracts on Spinach By Sreshta Vemula

Page 330: Predicting Post-Earnings Stock Performance in Small-Cap Markets: A Machine Learning Approach to "Buy the Rumor, Sell the News" By Mason Kozloff

Page 347: Alkali Metal Anodes for Rechargeable Batteries By Tim Zeng

Page 354: Climate Governance in Denmark and Australia: A Comparative Analysis and Lessons for Climate Action By Elizabeth Townsley

Page 365: Structural Causes and Policy Responses to Teacher Psychological Distress in South Korea By Taeyeon Kim

Page 374: The Effect of Financial Inclusion on Social, Economic, Political, and Environmental Indicators By Ardashir Prasetyo

Page 388: The Role of the Right Triangle in Engineering, Construction, and Architecture: A Study of Theory, History, and Practical Application By Timurkhan Abdullaev

Page 402: The Correlation Between Environmental Conditions, Socioeconomic Vulnerability, and Health Outcomes in the Delridge Community, West Seattle, WA By Lucy Powell

Page 411: Editing the Code of Life: CRISPR Technologies and the Future of Precision Medicine By Esther Zhou[1]*

Page 421: Democracy, Human Rights, and Women's Political Representation as Pathways to Better Living Conditions After Independence By Neel Venkatesan

Page 438: Art Therapy as a Non-Pharmacological Intervention: A Creative Bridge between Psychological and Neurocognitive Recovery By Sofia Neubauer[1]

Page 452: Donation Transparency and Consumer Skepticism in response to Cause-Related Marketing By Misha Cherukumilli

Page 467: -2.00 D: Rethinking ADHD as a Friction Between Mind and World By Lila Kaya

Page 474: A Comparative Study of Customer Satisfaction and Consumption Behavior Across Three Shopping Centers in Shanghai: Evidence from a Questionnaire Survey By Kaiqi Li

Page 488: Evaluating AI Trading Bots: A Comparative Study of Reinforcement-Learning Strategies By Navya Bhandari

Page 509: Democratic Backsliding By Malin Mehfar

Page 516: Simulating Stability: How Period Ratio Complexity Affects Orbital Resonance Lock in a Computational N-Body Model By Aarib Mahd Syed

Page 526: Play Vs. Academics in Early Childhood By Isabel Hershkin

Page 529: Is It Ever Wrong to Do the Right Thing for the Wrong Reason?: Morality Through the Kantian Perspective By Olivia Peng

Page 534: From Disparity to Equity: Understanding the Racial Wealth Gap and the Future of African American Wealth By Braeland Moten

Page 545: Mutation Directed Therapies in Pediatric and Adult Gliomas By Ekaterine Maghlakelidze and Mikheili Sharvashidze

Page 554: Delta Machine Learning for Ceiling Temperature Prediction: Bridging Computation and Experiment in Circular Polymer Design By Situo Xing

Page 567: Developing Sustainable Growth Policies: A 2035 Policy Transfer Model for Loudoun County and Kaiserslautern By Sarina Virmani

The Legacy of Riot Grrrl: Musician Career Evolution and Contemporary Perspectives
By Fern Preppernau

Abstract

In the early 1990s, the subcultural movement known as riot grrrl originated in the Pacific Northwest. Riot grrrl was an activist, feminist-oriented culture made up of primarily young women and girls, and it was closely intertwined with local punk scenes, with many key players of riot grrrl being musicians. After the controversial deterioration of the subculture near the end of the decade, many riot grrrl musicians continued to have lucrative musical careers continuing later, even into the present day. In this study, a content analysis was conducted of four former riot grrrl musicians' later published works, including autobiographical books, recorded interviews, and music. Analysis aimed to evaluate the amount of influence that riot grrrl had on these musicians' later careers, and how much similarity there was between riot grrrl-era work and later work. Results indicated that these musicians generally looked back at riot grrrl as complicated, having had both positive and negative impacts. Similarities in the themes addressed in music were also found, as well as significant specific evolutions in the discussion of topics like politics and gender roles. This research contributes to the scholarly conversation surrounding riot grrrl by offering a nuanced yet primarily positive perspective on the movement, emphasizing its lasting contributions to musical history and culture.

Introduction

What do you get when you combine the defiance of youth subculture, the politics of feminist activism, and the creative artistry of the punk scene? That blend is something close to a descriptor of the underground feminist punk movement called riot grrrl. In the early to mid 1990s, riot grrrl developed as a response to increasing political and social oppression faced by young women. One important aspect of riot grrrl was the creation of music, and as riot grrrl gained notoriety, its musicians and bands quickly took center focus in the public eye. However, in the late 1990s, they began to distance themselves from the movement, which correlated with its decline in influence.

Multiple musicians who started out as part of riot grrrl went on to have larger musical careers. Their careers for the most part have not been studied by scholars in the context of their riot grrrl origins, and this matters because it's a specific perspective of riot grrrl that hasn't been told. This study will consider those musicians and focus on what their relationship to the riot grrrl movement has continued to be throughout their lives.

Existing research has examined riot grrrl in hindsight through various lenses, mainly focusing on how the movement formed, actions it took as it existed (Marcus), or how it was treated by external factors (Mantila). Research has also considered riot grrrl bands' work within the timespan of the movement (Hasson). However, no research has been done that specifically examined the effects of the movement's end on riot grrrl musicians, or on their lives and work since. Understanding riot grrrl's true importance requires consideration of the complete story of

riot grrrl, including that of its musicians. This study also has the benefit of a greater period of hindsight than others, which allows for more nuanced reflection in order to better situate riot grrrl within history.

This study examines former riot grrrl musicians' relationships to riot grrrl after their time involved with the movement. Specifically, it asks: To what extent have former riot grrrl musicians maintained connection to the riot grrrl movement, even after their direct involvement ended? To address this question, this study uses a methodology of content analysis in two phases, analyzing the work of former riot grrrl musicians.

This research has significant implications for historical discourse about the riot grrrl movement. Riot grrrl is still perceived by some media and scholars as having been entirely ineffectual, while also being idealized by others to a potentially unrealistic degree. An increased understanding of musicians' later perspectives will help scholars continue to evaluate the movement's lasting impacts and overall importance.

Overview and Role of Music

Riot grrrl was an influential yet controversial sector of 1990s feminism that incorporated aspects of punk subculture. It was initially born out of alternative scenes specifically in Olympia, Washington, and also began to grow and develop in Washington DC shortly after. The movement had two key intertwined aspects: connection and creation. Riot grrrl chapters held meetings for women and girls to come together and build community over shared experiences. Connection was also built through the creation of art, like zines and music (Schilt 2003 p. 6-7). Zines are self-published, magazine-like publications with limited distribution (Tinio). They allowed riot grrrls to talk about taboo subjects and spread information about feminism without the usage of mainstream channels (Schilt 2003 p. 6). Music similarly allowed riot grrrls to spread feminist information and build connections, and riot grrrl-associated music and musicians will be the focus of this study.

It's important to note that riot grrrl was more than its music. Leading scholar Sara Marcus describes how by the late 1990s, riot grrrl was commonly thought of as having been a music scene or subgenre. While music did play a major role, its original purpose was to convey the movement's sociopolitical messages and to facilitate connection between community members. But although music was intended to further a larger purpose, bands emerged from the scene whose primary goals were to make music for its own sake (Schilt 2004 p. 122). Many of these bands gained popularity and influence. Scholarly consensus is that the three most significant bands of the riot grrrl scene were Bikini Kill, Bratmobile, and Heavens to Betsy (Marcus; Mantila; Hasson).

In her 2023 thesis, Vida Hasson briefly characterizes the nature of each band's formation, work, and ideology. Bikini Kill was founded very early on by Kathleen Hanna, intended to convey political messages after she was told that music was a powerful way to do so. This is evident in their work, as the gender politics of sexual violence and patriarchal systems were often addressed. Alongside Hanna were Tobi Vail, Kathi Wilcox, and Billy Karren. Bratmobile was

formed by Allison Wolfe and Molly Neuman, who were inspired by Kathleen Hanna and Tobi Vail to start a band. Erin Smith later completed the group. Their work often utilized a playful and childish tone, frequently referencing girlhood. Heavens to Betsy was made up of Corin Tucker and Tracy Sawyer, and was formed in 1991, the last of the three to form. Their work took a comparatively more intersectional perspective than their contemporaries.

Musicians' Later Experiences

The experiences of riot grrrl musicians were different than those of other riot grrrls, and this can be seen by considering media portrayals. During riot grrrl's prime in the 1990s, it was frequently misrepresented, rudely critiqued, and overall treated poorly by mainstream media. In her masters' thesis, Susanna Mantila establishes how media misinformation "sought almost solely to discredit the ideas of riot grrrls" (p. 6). For example, *Newsweek* accompanied an article about riot grrrl with a purchased picture of Bikini Kill lead singer Kathleen Hanna and her friends in bikinis in an attempt to "discredit their feminism" (Schilt 2003 p. 9).

Mantila writes that many outside writers "became preoccupied with the music created within [riot grrrl]" (p. 9), and ignored its other aspects. As such, many other media portrayals were by music-centered magazines or journals. This fixation on music meant fixation on those creating it, and focus was often shifted away from music to instead focus on musicians' personal lives in an invasive and harmful way. A 1996 article by *Spin* magazine about the riot grrrl associated rock band Sleater-Kinney nonconsensually outed band members Carrie Brownstein and Corin Tucker as queer women (Babovic).

It can be seen that the harm media did to riot grrrl as a whole also harmed individual musicians personally, and the effects were demonstrable. By around 1996, Bratmobile and Heavens to Betsy had broken up and Bikini Kill was on a long-term unofficial hiatus. The commonly accepted belief was that the movement essentially ended then, once musicians stepped away from it. While local riot grrrl chapters and communities did continue to develop in the years afterwards, riot grrrl had majorly changed and the creation of music was no longer as integral to the movement (Marcus). This was in large part due to the harms experienced by riot grrrl musicians.

Later Scholarly Conversation

Riot grrrl had fallen out of the public eye by the late 90s, but by about 2010, it began to be reconsidered, this time through a historical lens (Keenan). With this transition has come conflict over how that history should be told. Some historical narratives have been criticized as too idealized, being nostalgia-led and sacrificing nuance or accuracy because of it (Keenan p. 46). For example, Sara Marcus's *Girls to the Front* has been described as, in its romanticization of riot grrrl, narrating the stories and perspectives of women of color as obstacles or painful memories for the white core movement (Nguyen p. 188). Other historical essays offer sophisticated critiques of the movement. In her landmark essay "Riot Grrrl, Race, and Revival", Mimi Thy Nguyen criticizes riot grrrl's treatment of women of color and other marginalized

identities. Additionally, Charlotte Briggs points out the elitism and exclusivity that was present. Overall, recent studies of riot grrrl in the scholarly sphere have shifted to a critical lens, drawing attention to the movement's flaws.

Due to the harms they experienced within the movement as well as the critical perspectives now adopted by scholars, it seemed probable that musicians would perceive riot grrrl in a primarily negative light, and that they would have continued to distance themselves from riot grrrl as their careers went on.

Research Gap

The gap that this study addresses combines a stakeholder and a time gap. Research has explored the experiences of musicians within the riot grrrl movement up until they ceased involvement, but has not closely followed their evolution after, and no research exists that explores their recent perspectives. Studies focusing on hindsight perspectives of the movement have not looked specifically at musicians' perspectives, and conversely, research that focuses on riot grrrl musicians has not included their later careers.

This study fills that gap by exploring some of the most notable involved musicians' later relationships to riot grrrl. The study asks: To what extent have former riot grrrl musicians maintained connection to the riot grrrl movement, even after their direct involvement ended? Connection here refers to both explicit opinions of riot grrrl and implicit alignment with riot grrrl ideology.

To explore both explicit and implicit connection, this study will use a dual-phase content analysis. It will analyze two types of works created by focus musicians: autobiographical materials, and music. Autobiographical materials will reveal musicians' publicly expressed opinions on riot grrrl in hindsight— meaning explicitly stated connections. Lyrical analysis that focuses on ideals and values expressed within music, modeled after two studies by Hasson and Cederholm respectively, will reveal more subconscious connections: like whether or not lyrics still reflect riot grrrl ideals.

Methodology

Case Selection

To select cases from which to gather data, there were several factors to take into account. Firstly, this study chose to frame data selection by choosing four individual musicians as cases. Rather than focusing on individual bands as did Hasson, this study aimed to align with the specific research inquiry by focusing on individual people and their evolution over time.

The bands Bikini Kill, Bratmobile, and Heavens to Betsy are perceived as the most quintessential of the riot grrrl movement (Marcus; Mantila; Hasson), and the frontwomen of those bands (Kathleen Hanna, Allison Wolfe, and Corin Tucker respectively) each had further musical careers after the riot grrrl movement. Because of both their perceived influence and their longer-spanning careers, those three cases were selected as focus musicians for the study. The

fourth focus musician, Carrie Brownstein, was selected because of her work in the band Sleater-Kinney along with Corin Tucker. Sleater-Kinney is an example of a band that formed during the riot grrrl movement but evolved into a rock band over its career (Grønstad), and Carrie Brownstein was also part of an earlier riot grrrl band, Excuse 17, before Sleater-Kinney. For those reasons, musical evolution is apparent over the course of her career, which is why she was also selected. The study of these four musicians consisted of two phases, the first phase analyzing autobiographical (AB) materials of the focus musicians and the second phase analyzing lyrics of their music.

1st Phase: Autobiographical Content Analysis

Materials used were the autobiographies of Kathleen Hanna and Carrie Brownstein, a master's thesis published by Allison Wolfe, and Corin Tucker's interview with Fabi Reyna on the *Starting a Riot* podcast. These materials were selected because they met two criteria. First, autobiographical: something that "relates to events in the life of the person who has [created] it" (Collins Dictionary, s.v. "autobiographical"), and second, music-focused: centering on each person's musical career. Selecting materials where musicians reflected on their life history, centering on their musical careers, readily invited reflection on their opinions of riot grrrl overall.

This analysis was driven by a couple of guiding questions, primarily looking to discover the strengths and flaws of riot grrrl that each musician had perceived. Any songs mentioned by name were also noted in order to aid in data selection for the second phase of methodology.

Then, using inductive coding, data gathered from the AB materials for both strengths and flaws was coded into the themes that came up the most often. The results were graphed by how many out of the four musicians mentioned each theme, and topics mentioned by more than one musician were deemed most important for further analysis in order to focus on the most prevalent perspectives.

Analyzing this revealed what these musicians' current opinions looking back on the riot grrrl movement are, and what level of connection to it they say they maintain.

2nd Phase: Lyrical Content Analysis

Once the four focus musicians were selected, a chronological list was made for each musician of all the different bands they were in throughout their career. This was done in order to have a more complete picture of the change in their work over time.

Kathleen Hanna	Bikini Kill, Le Tigre, The Julie Ruin
Allison Wolfe	Bratmobile, Partyline, Sex Stains
Corin Tucker	Heavens to Betsy, Sleater-Kinney
Carrie Brownstein	Excuse 17, Sleater-Kinney

Figure 1: This table lists each band that each focus musician has been part of.

This study did not have the scope to analyze the entire discography of each band, so parameters had to be set. Each band's first album was analyzed, as well as the last album of the four bands with longer-spanning careers, as that most dramatically showed the changes in their songs over time. Other songs analyzed were each band's five most popular songs on Spotify, as well as any songs mentioned by name by musicians in the materials analyzed in the first phase of methodology. The wide range of selection criteria was used in order to gather a wide sample of each musician's songs. The second phase of methodology was analysis of the lyrics of these selected songs, then inductive coding for lyrical topics and themes.

Lyrical analysis by Hasson was used as a model for my coding in order to determine what themes aligned with riot grrrl. Hasson qualitatively analyzed the lyrics of 1989-1997 music by Bikini Kill, Bratmobile, and Heavens to Betsy, and each of my coded themes aligns with a topic her study mentions.

In analyzing the data, songs were divided into three time periods: 1990-1996 (P1) being true riot grrrl, 1997-2009 (P2) being initially post-riot grrrl, and 2010-2025 (P3) as being after riot grrrl's transition into history, more present-day. 89 of the songs analyzed were from P1, 65 were from P2, and 37 were from P3. Because of that, this data is expressed not as a raw number of songs that mention a theme, but as a ratio of the number of mentions in the time period over the number of songs from the time period, in order to get a percentage of how many songs per time period mention each topic. Analyzing this revealed to what extent riot grrrl musicians' expression of riot grrrl ideals in their music has changed over time, demonstrating the level of subconscious connection to riot grrrl their songwriting has contained.

Ethical Considerations

There were very few ethical considerations with my research because my methodology did not use human subjects. All content analyzed was publicly available and is credited under Works Cited.

Results

1st Phase: AB Materials

Below is a graphical representation of musicians' perceived strengths and flaws, graphed by how many of the four focus musicians made mention of each topic. There were three different strengths and five different flaws that were mentioned by multiple musicians each, and those eight topics will be what this analysis focuses on.

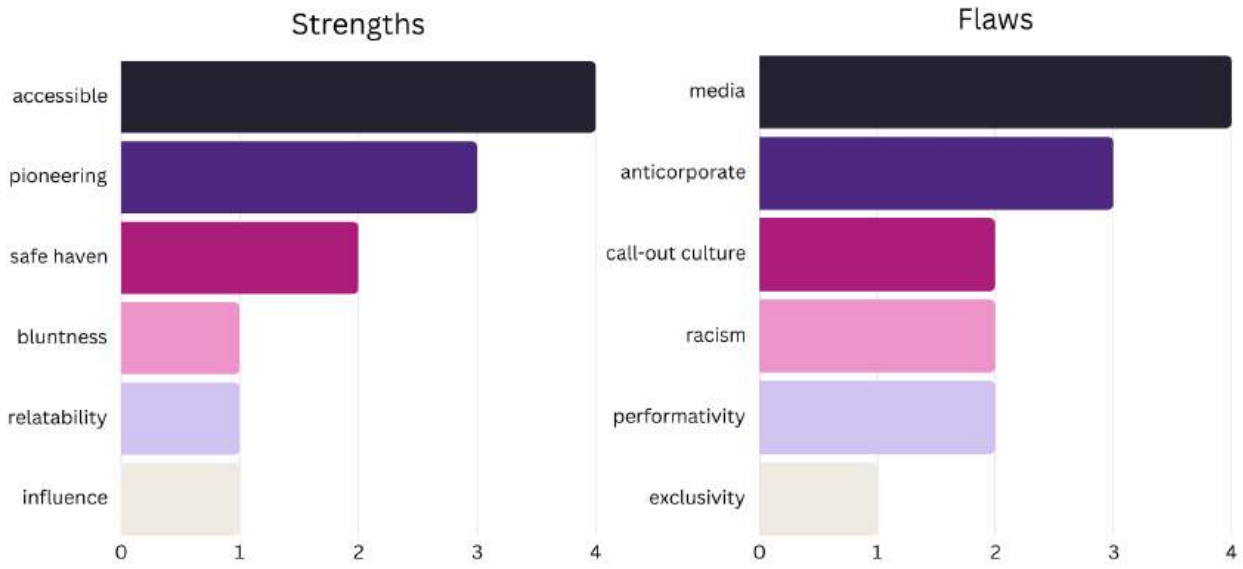


Figure 2: The graph above represents how many of the four focus musicians mentioned each strength and flaw of riot grrrl.

The musicians' most widely perceived strengths of riot grrrl were that it was **accessible** (4 musicians), **pioneering** (3 musicians), and a **safe haven** (2 musicians).

Accessibility refers to providing a venue in which people could learn about and participate in something in a positive and encouraging format. The data reflects the benefit of musical accessibility, for example: Carrie Brownstein's reflection on seeing a punk show as "...the first time [she] realized that music was playable, not just performable," (p. 8-9). Riot grrrl provided spaces to perform with few entry barriers, allowing aspiring musicians to learn by doing, even if their music initially sounded amateurish or unpolished.

Pioneering is defined as essentially coming first and paving the way for future examples to have an easier time doing the same thing. Corin Tucker reflected that now, and in part due to riot grrrl, there are "so many more doors that are open now for musicians", especially those who are female, queer, and trans ("Bonus Episode: Corin Tucker Interview", 22:11).

Safe haven refers to providing women and girls with a place in the music scene where they were not marginalized or threatened by men. Kathleen Hanna described wanting to do "everything in [her] power to make Bikini Kill shows a brief reprieve from sexism" (p. 148). One way this manifested was that, at the beginning of each show, Hanna would call for girls and women in the audience to come to the front, centering themselves in the space.

The most widely perceived harm to the movement was its interaction with **media** (4 musicians), followed by its **anticorporate** nature (3 musicians), followed by **callout culture** (2 musicians), **racism** (2 musicians), and **performativity** (2 musicians).

Media harm refers to mainstream media's coverage of the riot grrrl movement, which

was often condescending, caricatural, or entirely inaccurate. Kathleen Hanna specifically described how it led to the declaration of a "media blackout", a blanket refusal of riot grrrl members to speak to journalists.

Racism refers to the domination of the riot grrrl movement by white women and how the movement repeatedly marginalized the issues and concerns of women of color. This was mentioned by Kathleen Hanna, who admitted to herself having unwittingly been a contributor, recounting a failed experience of trying to run a workshop focused on intersectionality.

Anticorporate as a code for this data refers to the vehement anticapitalist or anti-"sellout" nature of the riot grrrl scene. Allison Wolfe described a negative response to the band Bikini Kill's request to be paid for playing shows: "some people thought that was too capitalist, but Bikini Kill was just squeaking by," (p. 28).

Callout culture means members of riot grrrl being torn down from within for making mistakes. Carrie Brownstein described being scared to act or participate fully in the riot grrrl scene for a time because of this, reflecting that "each step felt like a landmine" and how "the perimeters [of acceptability] felt unclear, almost like traps" (p. 99-100).

Performativity means losing sight of the politics in favor of the aesthetics. Allison Wolfe became frustrated with riot grrrl because she perceived that "various people [...] didn't seem interested in being politically progressive anymore," (p. 32).

2nd Phase: Lyrical Analysis

The inductive coding revealed six distinct riot grrrl-adjacent themes, each of which is discussed below.

Gender Roles - Addressing means songs that call out gender roles and expectations for women, for example expressing displeasure with a caretaker role, being seen as innocent, or being objectified.

Gender Roles - Defying means songs that, while not explicitly calling out gender roles, implicitly defy them via the speaker expressing 'unfeminine' behavior. These include songs that express anger, express sexual desire, or express a desire for gendered role reversal.

Female Empowerment means songs that seek to empower girls and women, for example the speaker describing themself as powerful, praising other women, or expressing power through activism or creation.

Politics means songs that refer to larger-scale American political issues or government policy, for example those that mention war, 9/11, or specific political figures.

Anticapitalism means songs that criticize commercialization, inequality, or issues in American society through an economic lens.

Music Scene Critique means songs that refer to the experience of being a musician or a female musician. These songs often call out issues, like racism, sexism, or rape, that musicians have experienced or witnessed in music scenes they've been involved with.

The presence of these themes by time period is shown here.

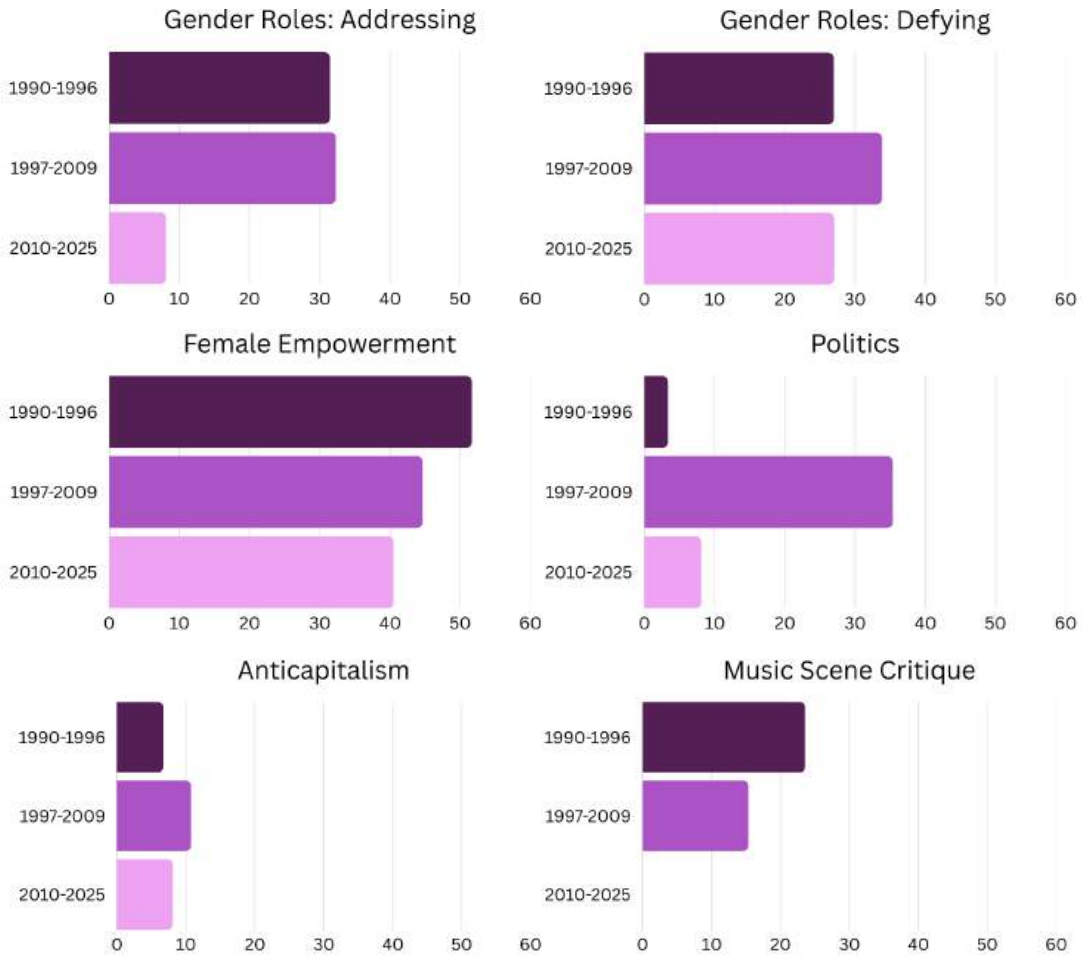


Figure 3: The graph above represents what percentage of songs in each time period mentioned each riot grrrl-associated topic.

Gender Roles: Addressing was mentioned in 31.4% of P1 songs, 23.6% of P2 songs, and 8.1% of P3 songs. Gender Roles: Defying was mentioned in 27.0% of P1 songs, 33.8% of P2 songs, and 27.0% of P3 songs. Female Empowerment was mentioned in 51.7% of P1 songs, 44.6% of P2 songs, and 40.5% of P3 songs. Politics was mentioned in 3.4% of P1 songs, 35.4% of P2 songs, and 8.1% of P3 songs. Anticapitalism was mentioned in 6.7% of P1 songs, 10.8% of P2 songs, and 8.1% of P3 songs. Music Scene Critique was mentioned in 23.6% of P1 songs, 15.4% of P2 songs, and 0% of P3 songs.

The data overall shows slightly more riot grrrl ideals in the initial post-riot grrrl period than in the true riot grrrl period. Female empowerment was the ideal most present overall, appearing in over 40% of songs from each time period, showing that the heart of riot grrrl stayed consistent in music. It was most common in true riot grrrl songs, such as Bikini Kill's "Double Dare Ya" singing "dare ya to do what you want / dare ya to be who you will". Music scene critique was also most common in true riot grrrl music, although it also had presence in P2, specifically Sleater-Kinney's 2000 album *All Hands on the Bad One*, which had multiple songs

responding to sexual assaults that occurred at the Woodstock ‘99 music festival.

Gender roles, both addressing and defying, were a common topic across all time periods. However, they appeared slightly more in the post-riot grrrl period than the true riot grrrl period, which contradicts the idea that the riot grrrl movement ended completely in the late 1990s. Another point of note is the difference between addressing versus defying gender roles within the present-day period. Only 8.1% of P3 songs directly addressed gender roles, while by contrast, 27% of P3 songs implicitly defied them. This is prominent in Sleater-Kinney’s 2024 album *Little Rope*. None of its 10 songs directly refer to gendered roles or expectations, but throughout the album the singer refers to herself with words like “wild”, “reckless”, and “untidy”, which are not traditionally feminine characteristics. This is a major contrast in approaches to the same topic, inviting conversation as to why.

The topics of politics and anticapitalism are each most present in P2, and these are also the two topics that don’t directly connect to gender. A common theme in songs from this time period was 9/11 and the subsequent war on terror, with major examples coming from Sleater-Kinney’s album *One Beat* (2001) and Bratmobile’s *Girls Get Busy* (2002). In the song “Far Away”, Corin Tucker sings “and the president hides / while working men rush in / and give their lives”, and in “Shop for America”, Allison Wolfe sings “I won’t fight your f****** war”, expressing disconnection and disassociation from the American government and military. These direct responses to national politics had a major presence starting in P2. Anticapitalism has a smaller presence throughout each time period, but nevertheless is most present in P2, with 10.8% of songs expressing anticapitalism.

Conclusion

This research examined the lasting legacy of the riot grrrl movement by looking at the opinions and creations of former riot grrrl musicians over time. Analysis showed that musicians had nuanced reflections on the movement, highlighting strengths of accessibility, pioneering, and safety, while also reflecting on media bias, extreme anticapitalism, and lack of intersectionality as flaws within the movement. This mix of strengths and flaws corroborates scholarly conversation by demonstrating nuance with a slight negative lean, with three widely perceived strengths and five widely perceived flaws.

For the most part, the individual strengths and flaws also align with existing research. The strength of musical **accessibility** was discussed by earlier scholars who interviewed musicians, such as Sara Marcus— implying that musical accessibility was perceived as a positive both at the time of the movement and later in hindsight. **Pioneering** aligns, as scholars perceive the riot grrrl movement as having been influential overall. This finding, however, complicates the findings of Elizabth Keenan who cautioned against nostalgia causing inflated perceived importance. Based on Keenan, the potential for nostalgia by scholars and musicians alike must be taken into account when evaluating riot grrrl’s so-called objective importance. **Safe haven** reflects Vida Hasson’s description of the male-dominated punk scene in the 80s and 90s as being both linguistically violent in song lyrics and shows being physically violent spaces. Riot grrrl’s

perceived safety in contrast to other punk scenes' violence is a logical continuation.

The major harm of the media on the riot grrrl movement has been discussed extensively by scholars such as Susanna Mantila, who studied how media misrepresentations served to discredit the politics of riot grrrl. **Racism** aligns with scholars like Mimi Thy Nguyen, who discussed racism within riot grrrl in her 2012 landmark critique "Riot Grrrl, Race, and Revival". **Callout culture** and **performativity** as flaws align with social dynamics discussed by Sara Marcus, who described some riot grrrl communities as having "a bit of a contest of oppression" (p. 178) and "a tone [...] that left many people feeling haunted and edgy, constantly on their guard against saying or doing the wrong thing" (p. 313).

The unique data point here is **anticorporate** being seen as a flaw, which stands out as a major contradiction of prior scholars. Wright and Jacques focused extensively on how commodification of the movement harmed it, but no scholars have considered the harms of the opposite. This may be a new finding because of this study's musician focus group. These four musicians all achieved exceptional fame and success, which could have contributed to being personally impacted by anticapitalist mindsets, such as being labeled "sellouts" or criticized for the success they achieved. This adds a new dimension to views of riot grrrl and other punk music scenes, shining a light on the potential dark side of DIY, anticapitalist rhetoric.

Lyrical analysis revealed how riot grrrl ideals influenced musicians' later work, showing a continuation of those ideals stretching beyond the 1990s, with their songs having equally strong or stronger ideals in the time initially after the movement.

Each lyrical theme code was developed based off of existing literature of riot grrrl ideals laid out by Rosenberg & Garofalo and Hasson. The strong presence of each of these codes in the riot grrrl time period aligns with the literature review. However, four of the six themes had an equal or greater presence in P2, clashing with the idea referenced by Schilt and Marcus that riot grrrl was perceived as ending in 1995-1996.

Politics and anticapitalism were both significantly more common in P2 than P1. These findings add to the scholarly conversation by introducing a new phenomenon: increased mention of non-gender specific issues in the time period immediately after riot grrrl. This potentially reflects an expansion in the personal and political ideals of riot grrrl musicians.

The politics code specifically also creates conflict. Riot grrrl was a political movement, but by my definition, only 3.4% of songs in P1 contained political themes, while 35.4% in P2 were coded as political. However, the definitions of "politics" in both the coding and in scholarly conversation are murky. The umbrella of "political" can be used to describe references to government laws, policies, and actions, which is how it was used in this study. But "political" can also refer to the individual's experience as a political entity. Riot grrrl was described as political, but "focus[ing] more on the individual and the emotional than on marches, legislation, and public policy" (Rosenberg & Garofalo p. 810). Marcus corroborated this, describing riot grrrl songs "[applying] tactics of confessional songwriting to political issues" (p. 130), meaning that lyrics acted as if they described a personal experience as a way to make a political statement. While the study's code was flawed, all things considered, the measured increase from P1 to P2

shows that musicians expanded their ideals to focus less on individual experience and more on government, policy, and activism.

Overall, the results of lyrical analysis add to and somewhat complicate scholarly conversation. Four of the six "riot grrrl themes" of music— specifically politics, anticapitalism, and both addressing and defying gender roles— were more present in music in the second time period than the first. This complicates the previously established definition of riot grrrl, showing that while these musicians were no longer active in the movement, they still maintained and even increased their ideological messaging.

Considering both phases of methodology, this study demonstrates that former riot grrrl musicians continued to maintain significant connection to the movement's key values. This can be seen through reflections which emphasize the positives as well as the negatives, and through later music and lyrics that reflect similar ideals to riot grrrl.

It's important to note that this study had multiple limitations. Firstly was the potential for interpretive bias in coding. This was a qualitative study done by one researcher working alone, which means the interpretation of themes and topics had inherent subjectivity to it. This limits the accuracy of conclusions.

Additionally, the study's scope was narrow. It only focused on four key musicians, each of whom achieved exceptional fame and notoriety because of their work in the movement. This limits the generalizability of findings, as these perspectives and ideals cannot be assumed to be equally the case for every musician who originally participated in the riot grrrl movement.

The final limitation was that in the second phase of methodology, the sample size of songs from the third period was significantly smaller than that of the other two. This was due to the focus musicians simply not having as much music from the more recent period, but that still presents a limitation to the accuracy of the statistics. Furthermore, some of the songs from that third period did not have transcripts of the lyrics available, so analysis of the words and themes of the lyrics was more difficult than the others and themes may have been missed. These factors limit the accuracy of the analysis of P3 music, meaning the conclusion that riot grrrl ideals were maintained by musicians is most accurate up until around 2010.

Despite the limitations, this study has implications for further research in related fields. It introduces two potentially new trends that were not considered by prior research. Firstly, the idea of anticapitalist ideology having negative effects on the riot grrrl movement, and secondly, the expansion of political ideals and actions as former riot grrrl members aged. Future scholars could conduct deeper and more targeted studies on either of these possibilities in order to determine their true significance if any. Research could also explore whether or not these two trends are visible with members of other subcultural movements.

Future research, specifically lyrical analysis, could also focus on younger, riot grrrl-inspired bands and musicians in order to determine to what extent riot grrrl ideals have continued on to younger generations of musicians. No scholars have yet studied this new generation, and exploring how much influence riot grrrl had on younger musicians could contribute further to the scholarly conversation surrounding riot grrrl's significance.

Works Cited

Babovic, Jovana. Dig Me Out. Bloomsbury Academic, 2016.

Briggs, Charlotte. ""Why Every Girl Isn't a Riot Grrrl": Feminism and the Punk Music of Bikini Kill in the Early 1990s." New Errands: The Undergraduate Journal of American Studies, vol. 3, no. 1, Oct. 2015, https://doi.org/10.59236/ne3159810.

Cederholm, Camilla. "Exploring traditional school song lyrics through content analysis." Responsible Research-Papers from the 4th Qualitative Research Conference, 2012, pp. 9-23. Academia.edu, www.academia.edu/1996877/Exploring_Traditional_School_Song_Lyrics_through_Content_Analysis.

"Definition of 'autobiographical'." Collins Dictionary, HarperCollins, 2020, www.collinsdictionary.com/us/dictionary/english/autobiographical.

Gronstad, Asbjorn. "Sleater-Kinney's Electric Feminism." American, British, & Canadian Studies Lucian Blaga University of Sibiu, vol. 44, no. 1, 2025, pp. 74-91, https://doi.org/10.59236/ne3159810.

Hasson, Vida. "Performing Grrrlhood: A Lyrical Analysis of Riot Grrrl Music." CMC Senior Theses, 2023, scholarship.claremont.edu/cgi/viewcontent.cgi?article=4273&context=cmc_theses.

Jacques, Alison. "Performing Grrrlhood: A Lyrical Analysis of Riot Grrrl Music." Canadian Woman Studies/les cahiers de la femme, vol. 21, no. 1, 1 Apr. 2001, pp. 46-50, cws.journals.yorku.ca/index.php/cws/article/view/6905.

Keenan, Elizabeth K. "Riot Grrrl: Nostalgia and Historiography." The Oxford Handbook of Punk Rock, edited by George McKay and Gina Arnold, Oxford Handbooks, 2025. Google Books, https://doi.org/10.1093/oxfordhb/9780190859565.013.37.

Mantila, Susanna. "Ugly Girls on Stage: Riot Grrrl Reflected through Misrepresentations." University of Vaasa, 2009, osuva.uwasa.fi/server/api/core/bitstreams/bd6bd999-c640-4eaa-be03-70717390d6ff/content.

Marcus, Sara. Girls to the Front: The True Story of the Riot Grrrl Revolution. HarperCollins, 2010.

Nguyen, Mimi Thi. "Riot grrrl, race, and revival." Women & Performance: a journal of feminist theory, vol. 22, no. 2-3, Nov. 2012, pp. 173-196. Are.na, www.are.na/block/7760165.

Rosenberg, Jessica, and Gitana Garofalo. "Riot Grrrl: Revolutions from Within." Signs, vol. 23, no. 3, 1998, pp. 809-841. JSTOR, www.jstor.org/stable/3175311.

Schilt, Kristen. "'A Little Too Ironic': The Appropriation and Packaging of Riot Grrrl Politics by Mainstream Female Musicians ." Popular Music and Society, vol. 26, no. 1, 2003, pp. 5-16, doi:10.1080/0300776032000076351.

Schilt, Kristen. "'Riot grrrl is...' Contestation over meaning in a music scene." Music scenes: Local, translocal, and virtual, 2004, pp. 115-130. Google Books, books.google.com/books?id=zrGa3vYOoZgC&pg=PA118#v=onepage&q&f=false.

Tinio, Jerilyn. "Loud on Paper Library Guide: What is a Zine?" Purdue Libraries, 2022, guides.lib.purdue.edu/loudonpaper/what-is-a-zine.

Wright, Lindsay. "Do-it-yourself girl power: An examination of the Riot Grrrl subculture." James Madison Undergraduate Research Journal, vol. 3, no. 1, 2016, pp. 52-56, commons.lib.jmu.edu/cgi/viewcontent.cgi?article=1018&context=jmurj.

Data Materials

[various artists]. Kill Rock Stars. Kill Rock Stars, 1991. Spotify app.

Bikini Kill. Pussy Whipped. Kill Rock Stars, 1993. Spotify app.

Bikini Kill. Reject All American. Kill Rock Stars, 1996. Spotify app.

Bikini Kill. Revolution Girl Style Now. 2015. Spotify app. [later self-released collection].

Bikini Kill. The First Two Records. Kill Rock Stars, 2014. Spotify app.

Bikini Kill. The Singles. Kill Rock Stars, 1998. Spotify app.

"Bonus Episode: Corin Tucker Interview." Guest Corin Tucker. Starting a Riot, hosted by Fabi Reyna, podcast ed., season 1, episode 8, Oregon Public Broadcasting, 27 July 2023, www.opb.org/article/2023/07/27/starting-a-riot-bonus-episode-1-corin-tucker-interview/.

Bratmobile. Girls Get Busy. Lookout! Records, 2002. Spotify app.

Bratmobile. Ladies, Women, and Girls. Lookout! Records, 2000. Spotify app.

Bratmobile. Pottymouth. Kill Rock Stars, 1993. Spotify app.

Bratmobile. The Real Janelle. Kill Rock Stars, 1994. Spotify app.

Brownstein, Carrie. Hunger Makes Me a Modern Girl: A Memoir. Riverhead Books, 2016.

Excuse 17. Such Friends Are Dangerous. Kill Rock Stars, 1995. Spotify app.

Hanna, Kathleen. Rebel Girl. HarperCollins, 2024.

Heavens to Betsy. Calculated. Kill Rock Stars, 1994. Spotify app.

Heavens to Betsy. These Monsters Are Real. Kill Rock Stars, 2018. Spotify app.

The Julie Ruin. Run Fast. Dischord Records, 2013. Spotify app.

Le Tigre. Le Tigre. Mr. Lady Records, 1999. Spotify app.

Partyline. Girls With Glasses. Retard Disco, 2005. Spotify app.

Partyline. Zombie Terrorist. Retard Disco, 2006. Spotify app.

Sex Stains. Sex Stains. Don Giovanni Records, 2016. Spotify app.

Sleater-Kinney. All Hands on the Bad One. Kill Rock Stars, 2000. Spotify app.

Sleater-Kinney. Call the Doctor. Chainsaw Records, 1996. Spotify app.

Sleater-Kinney. Dig Me Out. Kill Rock Stars, 1997. Spotify app.

Sleater-Kinney. Little Rope. Loma Vista, 2024. Spotify app.

Sleater-Kinney. No Cities to Love. Sub Pop, 2015. Spotify app.

Sleater-Kinney. One Beat. Kill Rock Stars, 2002. Spotify app.

Sleater-Kinney. Sleater-Kinney. Chainsaw Records, 1995. Spotify app.

Sleater-Kinney. The Hot Rock. Kill Rock Stars, 1999. Spotify app.

Sleater-Kinney. The Woods. Sub Pop, 2005. Spotify app.

Wolfe, Allison. "Girl Germs, No Returns: A Bratmobile Oral History." University of Southern California Masters Theses, May 2017.

The Paradox of Underrepresentation in Female Leadership By Yeryeong Chun

Abstract

Despite increased female participation in the workforce, women remain significantly underrepresented in corporate leadership roles compared with men. Existing studies attribute this gap to systemic bias in the corporate world. However, previous research lacks a detailed explanation of the gendered characteristics behind this systemic bias, or of how early these perceptions of leadership are formed. This paper aims to provide a deeper understanding of existing gender biases by analyzing the students' perceptions of gender leadership traits. This research implements a survey research design that collects data from high school students in Vietnam. A literature review was then synthesized with the questionnaire findings to compare and contrast the youth perceptions against the real-world corporate metrics. Ultimately, the literature review corroborated the data analysis, supporting the conclusions drawn about the limitations for females in higher corporate leadership positions. A paradox was revealed as female leaders were rated as more effective leaders overall with organizational and operational skills, yet still remain underrepresented in the real world. Hence, this suggests that the disparity may be caused by the confidence gap and gender stereotypes rather than by women's actual competence.

Introduction

The disparity between female and male CEOs in corporate leadership has been a persistent topic of discussion in our global economy. The latest CEO Turnover Index found that, even with more women taking CEO roles than ever before, only 15% of new CEOs appointed globally in 2023 were women (Meynell). This severe underrepresentation of women is presented as a paradox that can act as a disadvantage to the economy, as research has shown that a better gender mix among the corporate leadership roles results in higher return on equity, valuations, better stock performance, and many more benefits (Lien and Thuy 89). Yet, the real-world corporate metrics show that the firms fail to close this gender gap in executive leadership. This raises a critical question: why does a paradox exist between the proven competence of the female leaders and their representation in executive positions?

Existing literature has traditionally focused on the structural and functional barriers for female leaders. However, these studies often overlook the psychological and cognitive factors that fuel this disparity. There is a prominent research gap in existing literature on how the cognitive leadership traits can contribute to the female underrepresentation in corporate leadership, such as gendered biases and the confidence gap (Abrams). Moreover, most studies fail to identify how these perceptions are formed in the early stages of people's minds. It is critical to examine how cognitive gender biases are formed before people enter the working environment, and how they can influence the patterns in the workforce. This study aims to address the gaps by surveying high school students in Vietnam and comparing their perceptions with real-world metrics. This research offers an insight into the cognitive reasoning behind the

gender gap that prior research has not approached. Hence, this paper examines the research question, 'To what extent do the perceptions of Vietnamese high school students on gender and leadership align with the gender imbalance in corporate leadership roles?'

Literature Review

Underrepresentation of Female Leadership

Existing literature identifies the persistent underrepresentation of female executives in the senior executive field. In the 500 S&P companies, females make up 45% of the population of the workforce, but only hold 25% of the executive leadership and senior-level managerial roles (Catalyst). The top of the workforce presents an even lower rate of gender parity, as females account for 11.2% of CEO roles (Catalyst). This further supports how the gender leadership disparity is not only present in the workforce, but especially restricted at the top-tier, senior executive positions.

This shortfall in female CEOs is evident in Vietnam as well, which is relevant for this study because the survey takes place among students in Vietnam. The Grant Thornton International Business Report, "Women in Business 2025", presents that in contrast to the rise in the occupation of female senior leadership positions, the share of women among CEOs in Vietnam declined from 22.5% in 2024 to 12.5% in 2025 (Grant Thornton). Similarly, Tran and Nguyen's 2020 study indicates that although positive perceptions of female leadership have been observed, there has been a lack of opportunities provided for women (Tran and Nguyen). This research shows an unconscious preference towards the male population and raises a question about the perceived legitimacy of female leaders in the context of Vietnam.

Barriers for Women in Corporate Leadership

Past literature addresses the structural barriers to women that contribute to the lack of leadership opportunities for females. In Vietnam, access to financial support and concessional loans for women remains significantly lower compared to men. Statistics show that only 37% of the female-owned businesses are able to obtain loans, while the figure is 47% for male-owned firms (Freeman). Within the companies, women also receive less of the sponsorship and support that drives high-level promotion. According to the 'Women in the Workplace' report, only half of the companies in their study prioritized female career advancement (McKinsey & Company). There is a significant lack of sponsorship provided and manager advocacy, which is critical for their promotion and leadership advancement. Even when entry-level women do have a sponsor, they are observed to be promoted at a slower rate than men (Field et al.).

Competency of Female Leaders

Research by The Leadership Circle and Zenger/Folkman found that female leaders rate higher than men in many key leadership competencies, such as taking initiative, driving for results, and self-development (Zenger and Folkman). Specifically, a study found that female

leaders were rated higher than male leaders across every management level and age group (Kruse).

Female leaders are also shown to help advance the firm financially. The 'Ready-Now Leaders Report' from the Conference Board showed that "organizations with at least 30% women in leadership roles are 12 times more likely to be in the top 20% for financial performance" (Kruse). In addition, Khan and Vieito's study conducted in 2013 illustrates that firms with female CEOs demonstrated stronger financial outcomes, with lower leverage and capital expenditure (W. A. Khan and Vieito 55). Meanwhile, they gained higher earnings, returns on assets, and tangibility (W. A. Khan and Vieito 55).

By combining this data, a paradox can be identified in the corporate culture. Although females demonstrate strong leadership capabilities and are proven to bring a positive impact to the company's performance, female leaders remain largely excluded from the top corporate positions. This presents a strong contradiction about how the underrepresentation of female leadership coexists with clear evidence of their competence in their leadership roles. This study examines the perceptions of high school students in Vietnam to address the gap in existing research. This study hypothesized that the respondents would perceive females as more capable than males, presenting a deviation from real-world metrics, and that this deviation can be attributed to a difference in leadership style between the two genders.

Methodology

A mixed-methods study was conducted in the form of an online questionnaire, allowing for the analysis of both quantitative and qualitative data. The purpose of this questionnaire was to examine whether the perceptions of female leadership among students in Vietnam align with the existing metrics of the number of female executive leaders. This method aimed to ultimately assess whether these perceptions reflect the broader pattern of the underrepresentation of females in corporate leadership.

Measures

The study was conducted in the form of an online questionnaire as a Google Form. The questionnaire was newly designed and customized for this specific research. The questions in the survey aimed to evaluate the students' perceptions of the different genders and their leadership effectiveness, with questions such as 'What do you think are the strengths and traits of female leaders?', as well as how they perceive the different leadership characteristics for each gender. The full questionnaire is available in the Appendix. Although the questionnaire is not based on a previously used measure, some questions were derived from an established leadership theory about transformational and transactional leadership styles. This framework allowed for a clearer interpretation and analysis of the respondents' choices on the different characteristics associated with gender. Furthermore, the questionnaire was designed so that the questions are unbiased, close-ended, provide an opt-out, and that there is a balance in the numerical scales.

Data Collection

The data collection sampling included high school students in Vietnam, aged between 14 and 18. The study collected 110 responses from around 15 different public and private high schools across different regions in Vietnam (e.g., Hanoi International School, Hoi An International School, Saigon South International School) through social media direct messenger. To adhere to the ethical codes, all respondents were provided with an informed consent form for the purpose and use of the research paper. The responses were fully anonymous to protect the respondents' privacy and promote honesty.

Data Analysis

The data were filtered with data cleaning to remove incomplete or invalid responses. Two responses were deleted since one had all blank responses and the other had irrelevant, humorous responses.

The analysis of the quantitative data was carried out through different statistical methods. For the 'Which gender is more efficient at' questions, the responses were stratified by the gender of the respondent. Then, a cross-tabulation was run on the two variables — the gender of the respondent and their response. This was to accurately identify the potential demographic variations, examining whether the two genders presented different views about their own or opposite sex, allowing for an in-depth approach to the gender perceptions.

Then, the standard deviation was calculated for the scale-based responses where the respondents were asked to rate the leadership efficacy of the two genders. This helped examine the level of consensus and cohesiveness among the responses. For the remaining question types, the responses were grouped accordingly and translated as pie charts and bar charts. Translating the responses into charts helps visualize the data, aiding in analyzing the pattern or the maximum and minimum values. For the qualitative responses obtained from the open-ended questions, a word frequency analysis and word clouds were used for a visual representation of the patterns.

Limitations to the Methodology

One limitation of this research is that the sampling included the researcher's personal social media connections rather than a broad school-wide distribution, which may introduce selection bias. However, it was ensured that the sample size included students from different demographic and cultural backgrounds, as well as a variance in gender and nationality. Additionally, a high proportion of respondents selected the neutral option in their responses, which limits the extent of data that this survey could capture and hinders the representation of the whole population in the analysis. Furthermore, the findings represent only the perceptions of Vietnamese high school students due to the survey's sampling. Therefore, it is difficult to make a generalization on the perception trends for the whole world population. The students' perceptions of corporate world leadership are also only theoretical and don't include any personal experience, which indicates that their views might differ from those of the people in the actual workforce. However, an analysis of secondary data shows that the discussion of the study is not

only applicable to Vietnam, but a prevalent phenomenon. Moreover, for the data collection, data cleaning allowed for the removal of 2 invalid responses, leading to the figures yielding 108 valid responses. The number of responses also varied per question and figure due to some respondents skipping certain questions.

Results

The following section provides the results of the questionnaire.

Table 1: Identified Gender of the Leader by Organization/ Event Type

Question	Male leader	Female leader	I don't know
Famous organization/brand	70 (65.4%)	22 (20.6%)	15 (14.0%)
Student organization	17 (15.7%)	81 (75.7%)	10 (9.3%)
Major event	36 (34.0%)	32 (30.8%)	35 (33.7%)

Table 2: Identified Gender of the Efficiency by Leadership Skills According to the Respondents' Experience

Leadership skill	Male	Female	I don't know
Logistics planning/organization	20 (18.7%)	74 (69.2%)	13 (12.1%)
Team leadership	36 (33.6%)	51 (47.7%)	20 (18.7%)
Problem-solving	33 (30.6%)	51 (47.2%)	24 (22.2%)

Table 3: The respondents' perception of which gender works more efficiently and successfully as a leader in association with the Leadership Factors

Leadership factor	Male leaders	Female leaders
Long-term vision	43 (40.2%)	64 (59.8%)
Supervision and accountability	24 (22.4%)	83 (77.6%)
Motivating others with inspiration	56 (52.3%)	51 (47.7%)
Logistics and planning	24 (22.2%)	84 (77.8%)
Innovation and risk-taking	92 (85.2%)	16 (14.8%)
Rewards management style	46 (42.6%)	62 (57.4%)

Table 4: Identification of Which Gender of the Leader the Respondent would Trust more in a Group of People

Category	Number of Responses (n)	Percentage (%)
Female	53	50.50%
I don't know	42	40.00%
Male	10	9.50%
Total	105	100.00%

Table 5: Percentage and Mean Efficacy Ratings by Gender

Rating (1=Very Inefficient, 5=Very Efficient)	Male Leaders: Count & %	Female Leaders: Count & %
1	0 (0.0%)	0 (0.0%)
2	6 (5.6%)	1 (0.9%)
3	41 (38.3%)	18 (16.8%)
4	48 (44.9%)	71 (66.4%)
5	12 (11.2%)	17 (15.9%)
Mean Rating	**3.62**	**3.97**

Table 6: Main Themes and Common Words Identified for Qualitative Responses

Focus Area	Main Themes Identified	Common Words/Phrases
Reasons for Leadership Trust	Personal experience, Neutrality, Organization, Confidence	"Depends on the person," "more organized," "not about gender."
Justification for Efficacy Ratings	Male: High vision, poor execution Female: High competence, cautious	Male: "All talk," "slacking off," Female: "too careful," "overthinking."
Strengths of Male Leaders	Vision, Risk-taking, Confidence, Authority	"Confident," "Risk-Taker," "decisive"
Strengths of Female Leaders	Organization, Empathy, Communication	"Organized," "empathy"

Data Analysis and Processing

Stratification

Stratification was carried out by running a cross-tabulation on Google spreadsheets for the gender of the respondent. For the question, *'In your experience, which gender is more*

efficient at Logistics planning/organization?', the cross tabulation table showed the following results of the number and the percentage of each gender who responded 'Female' or 'Male'.

Table 7: Cross-tabulation of the respondent's gender with the respondent's perception of efficiency in logistics planning

Which gender is more efficient at logistics planning/organization?	**Female Respondents (n=61)**	**Male respondents (n=39)**	**Prefer Not to Say (n=7)**	**Total number of respondents (n=107)**
Female	41 (67.2%)	29 (74.4%)	4 (57.1%)	74 (69.2%)
Male	8 (13.1%)	4 (10.3%)	1 (14.3%)	13 (12.1%)
I don't know	12 (19.7%)	6 (15.4%)	2 (28.6%)	20 (18.7%)
Total	61 (100%)	39 (100%)	7 (100%)	107(100%)

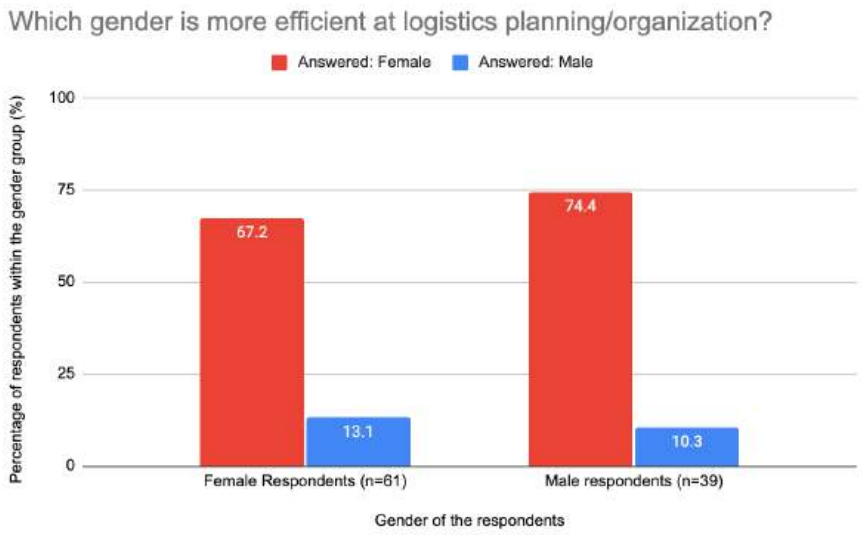


Fig. 1: A bar graph of the respondents' gender and their perception of efficiency in logistics planning

The cross-tabulation for this question revealed that, surprisingly, more male respondents stated that females are more efficient than males, compared to the number of female respondents who stated that. This differed with 74.4% of male respondents answering "Female," and 67.2% of female respondents answering "Female." This shows that the student respondents did not contain bias towards their own gender, and that female leaders are widely accepted as efficient at logistics planning and organization. This data conveys that the association of women with logistical skills is a widely shared belief that

transcends gender divisions. However, this paper later identifies that this consensus of strongly linking females with their operational competence is what traps them in the lower positions in the corporate firms.

Table 8: *Cross-tabulation of the respondent's gender with the respondent's perception of the leader's efficiency in team leadership*

Which gender is more efficient at team leadership?	**Female Respondents (n=62)**	**Male respondents (n=38)**	**Prefer Not to Say (n=7)**	**Total number of respondents (n=107)**
Female	37 (59.7%)	12 (31.6%)	2 (28.6%)	51 (47.7%)
Male	13 (21.0%)	19 (50.0%)	4 (57.1%)	36 (33.6%)
I don't know	12 (19.4%)	7 (18.4%)	1 (14.3%)	20 (18.7%)
Total	62 (100%)	38 (100%)	7 (100%)	107 (100%)

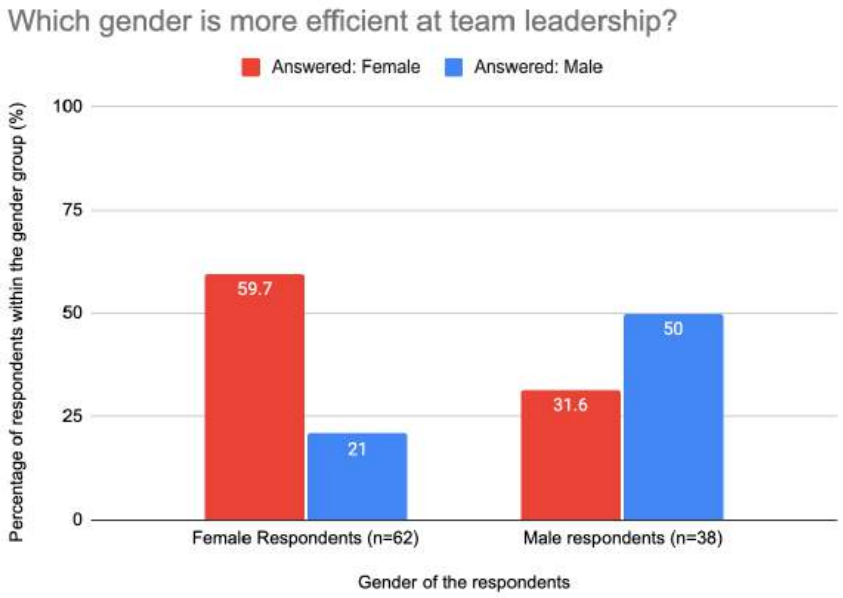


Fig. 2: A bar graph of the respondents' gender and their perception of the leader's efficiency in team leadership

For the question, 'In your experience, which gender is more efficient at Team Leadership?', the responses varied for each gender of the respondent. While females stated that 'females' are much more capable in team leadership with 59.7%, the males selected 'males' to be superior in team leadership with 50%. This can be compared to the data of logistics and planning in Figure 1, which shows that the perception of female competency is decreased when it is framed as 'leadership' rather than logistical planning. Specifically, while many male respondents acknowledged female superiority in logistics, half of them still associated 'Team

Leadership' with male figures. This shows that there is a clear disparity between how they perceive operational competence – logistics, planning, execution – with the broad term of authoritative 'leadership.' Respondents draw a line between being able to do the functional work in a group compared to being the leader, and gender perceptions play a critical role when associating these traits. This connects with how many females are limited to supporting, operational roles, and ultimately explains the real-world metrics of the number of female CEOs. This pattern aligns with Hoobler et al. 's finding that women often stay in the middle management positions while top roles seem unattainable for them (Hoobler et al. 703).

Standard Deviation Analysis

The average and the standard deviation for both male and female efficacy values were measured to determine the variance in respondent perceptions.

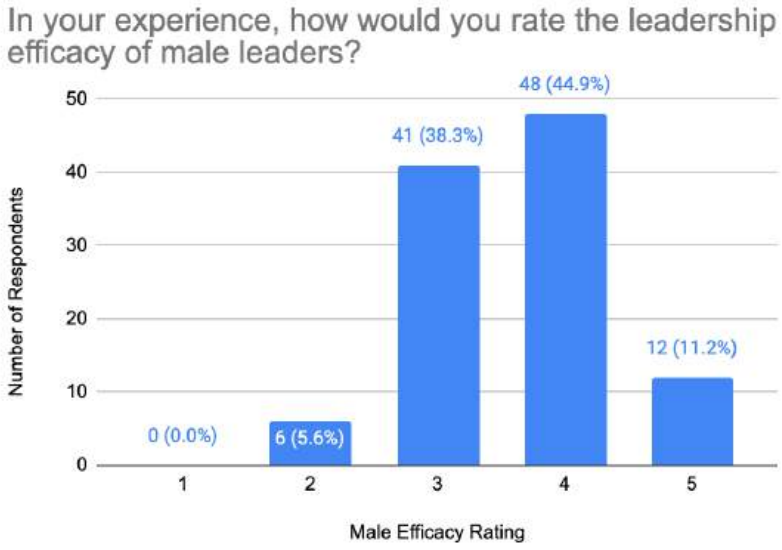


Fig. 3: A bar graph representing the male efficacy rating from 1 to 5

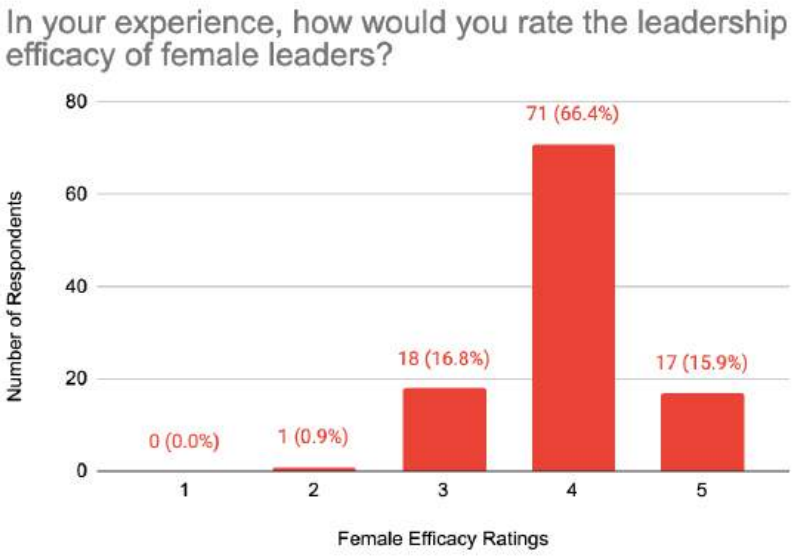


Fig. 4: A bar graph representing the female efficacy rating from 1 to 5

Table 9: The calculated standard deviation of leadership efficacy ratings for male and female

Gender	**Number of Responses**	**Mean**	**Standard Deviation**
Male	107	3.62	0.76
Female	107	3.97	0.61

The higher standard deviation for Figure 3 shows that the opinions were highly split on the leadership efficacy of male leaders. This may indicate that respondents had less consistent perceptions of male leadership. However, the lower standard deviation of 0.61 for female leadership shows that the high rating of female efficacy of 3.97 is a widely accepted and proposed opinion. This helps present a strong consensus that female leadership is consistently reliable. This enhances the paradox proposed in the study because the positive perception of female leadership is not concealed, and the barrier to women in the workforce is not about their actual competence.

Quantitative Frequency Analysis

The difference in quantitative frequency ratio was analyzed between the responses about successful brands and student organizations. The respondents were asked to state a successful brand that they know, then asked to identify the gender of the CEO of that brand. Similarly, they were asked to state a successful student organization that they know, then asked to identify the leader's gender. Accordingly, they were asked to provide an explanation for why they picked that answer.

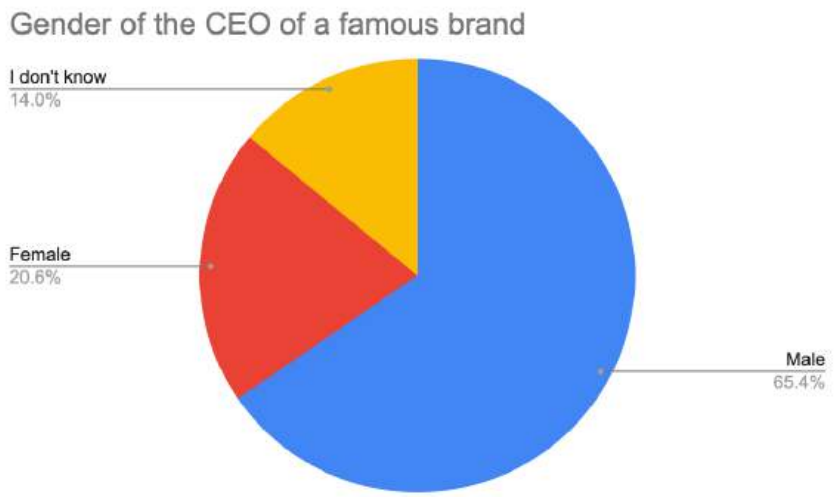


Fig. 5: A pie chart representing the respondents' identification of the gender of the CEO in a famous brand

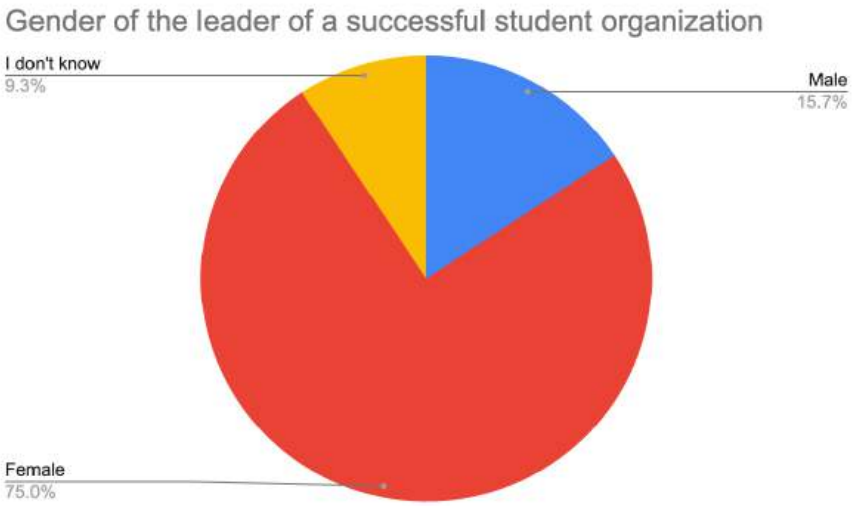


Fig. 6: A pie chart representing the respondents' identification of the gender of the leader in a student organization

Figure 5 and Figure 6 present opposite results, since 65.4% of respondents identified the CEO of a famous brand as a male leader, while 75.0% responded that the leader of a student organization is a female. One respondent offered a possible explanation for this disparity when asked "Why?" about their answer choice, as they stated: "Females tend to be more responsible and mature, especially while being teens." This respondent's reasoning suggests that women are likely to develop and mature faster, making it more likely for female students to be able to take on leadership roles and become successful. This indicates that while student organizations

prioritize functional leadership, famous firms value symbolic leadership. Student organizations require a lot of functional and transformational abilities, such as organizing and executing. The capability to carry this out for a student is rare, since most students are not taught how to carry out a big event or lead a big group of events individually. Women were observed to be more capable in organization and logistics, making them a more preferable candidate when leading a successful student organization.

On the other hand, the role of a CEO is often associated with traits such as vision, risk-taking, and firm representation rather than day-to-day operational management. The CEOs may tend to delegate these logistical responsibilities to subordinates, such as female employees. Therefore, it can be inferred from the data that the real world number of male leaders in the real world is attributable to how most CEOs are represented with visionary and risk-taking traits. This suggests that female workers are commonly associated with operational execution, so they eventually get delegated all the logistical work, while men take on the visionary face of the brand.

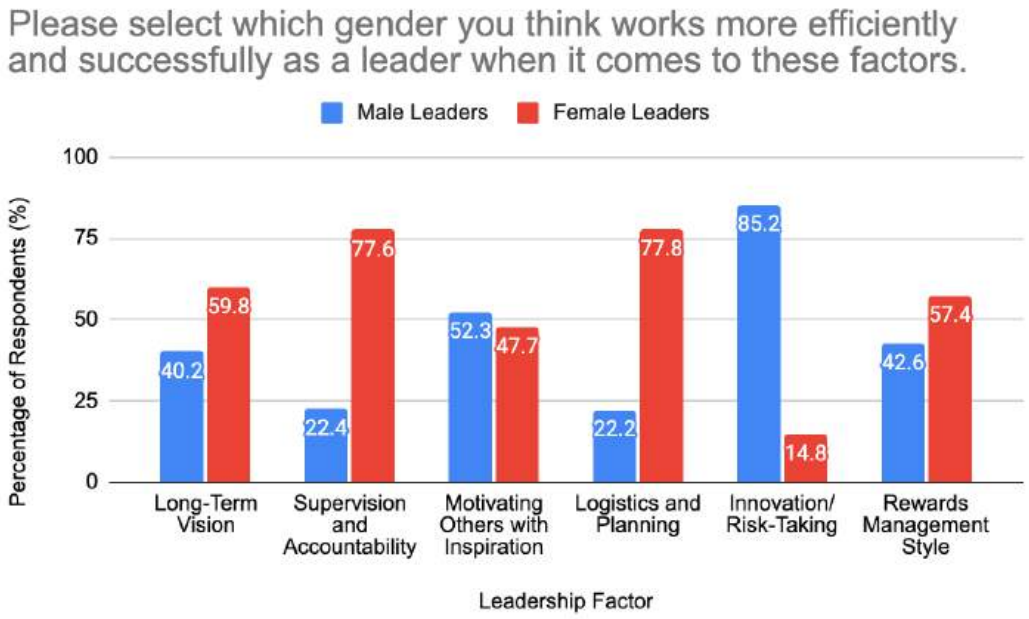


Fig. 7: Respondents' Gender Associations with Leadership Efficiency Across Different Leadership Traits

Figure 7 displays how respondents associate each of the six primary leadership styles with male or female leaders. A clear distinction can be analyzed from this figure, as most operational and executive leadership traits were associated with women, with 78.2% of respondents selecting females for 'Logistics and Planning' and 78.0% of respondents choosing females for 'Supervision and Accountability." In contrast, some categories were dominated by

votes for male leaders, such as 85.2% of respondents voted for men when it came to innovation and risk-taking, and 52.3% selected that male leaders 'Motivating others with inspiration.'

This shows how male leaders are often grouped with visionary, motivating traits that are considered transformational leadership. Interestingly, the high percentage of respondents for each category shows that these classifications mainly do not include a demographic bias of the respondents' gender, but rather that it is a universally shared perception. This is crucial because it is possible that transformational leadership, such as motivation and risk-taking, can be tied to attributes needed for top leadership, such as the CEO role. It can be assumed that since their logistical skills are so universally recognized and relied upon, female leaders are often pushed into the operational and planning roles instead, restricting their opportunities to be able to take on the bigger CEO roles. This aligns with the role congruency theory, where a top leader's characteristics link to agentic, visionary traits, while women are often praised for their operational competence, which is valued less for higher roles (Eagly and Karau). By identifying transformational and transactional leadership in terms of gender, it can be inferred that different types of leadership styles impact the real-world distribution of leadership positions.

Qualitative Analysis of Strengths

Qualitative data was collected through long-answer questions such as 'What do you think are the strengths and traits of male leaders?'



Fig. 8: Word Frequency Analysis of Leadership Traits Associated with Male Leaders



Fig. 9: Word Frequency Analysis of Leadership Traits Associated with Female Leaders

Figures 8 and 9 present word clouds where a bigger word size correlates with a higher frequency of the word in the survey responses. The first word cloud presents the frequency of words when asked for strengths associated with male leaders, and the second presents the strengths of female leaders.

For men, variations of "Confidence" or "Confident" were cited 19 times as a core strength, and 19.4% of respondents stated that male leaders are confident. In comparison, for women, the factor ‘confidence’ wasn’t indicated at all. This significant difference represents the confidence gap between men and women in their leadership positions. This is important because confidence is one of the crucial factors that is tied to successful leadership, and hence it can be inferred that the paradox is attributable to the confidence gap.

Variations of the word "Risk" (risk-taking, risk-taker, riskier) were stated 17 times in the male traits section. This contrasts with how people said that women actually don’t take risks. For females, they stated, “they don't take as much risk because sustainable practice is their priority”, which emphasizes their patient and well-constructed, logical nature instead of taking risks. And for the strengths for women, the word risk was mentioned only 2 times. This shows that the risk with men was stated 8.5 times more than it was for women.

Conversely, the words that can be associated with female leadership from these qualitative responses are "organization" or “empathy”. Variations of the word ‘organization’, such as “organize”, ‘organized’, and ‘organizing,’ appeared 23 times, making it the most frequently mentioned word when associated with female leadership. With words such as ‘detail’ and ‘planning’, over 30% of the qualitative responses were directly related to the functional execution carried out by female leadership.

When describing male leaders’ strengths, one response stated, “Communicative, humour, adaptable, outspoken, sociable that they can bring everyone together, they tend to be less shy than most females and they can bring everyone together.” Another respondent wrote, “Creating a good image that will put an impression on people, especially tall, or just guys who are

confident.” This portrays that male leaders are likely to show an outgoing and enthusiastic personality that makes them seem confident and well-heard. Interestingly, almost no responses emphasized the male leaders’ operational or logistical skills. Hence, the strengths of male leaders are often associated with how they present themselves externally to other people rather than the real operational work that they complete.

However, for females, the qualitative responses showed a different perspective. One person affirmed, “Decision-making, delegation, drive, dedication, fostering healthy working environments.” Many responses focused on how the female leaders improve the internal team setting, which is crucial for the operation of a firm. Another person asserted, “Well organized and has proper management.” This reflects that females are more acknowledged for their abilities in maintenance and execution within an organization rather than external representation, which can help explain the paradox of the real-world metrics.

Qualitative Analysis of Weaknesses

Several weaknesses were identified for the males from the analysis of the section ‘Why?’ One response asserted, “They make grandeur promises... Despite that, the work is actually harder than it seems so then they tend to slack off." This highlights the gap between how male leaders present themselves and their legitimate capabilities. This further emphasizes why females are so crucial to fill in this flaw in corporate firms.

Another trait is the lack of empathy. One response stated, “I think male leaders are able to be efficient, however a lot of them lack the empathy to see some workers as humans.” This contrasts with how ‘empathy’ was one of the most frequent strengths identified for women. This contradiction suggests that the barrier that women face in the corporate world can possibly be associated with this specific strength, as empathy contradicts the dominant and tough stereotype of an executive leader. This links to Schein’s study, which states that “successful middle managers are perceived to possess characteristics, attitudes, and temperaments more commonly ascribed to men in general than to women in general” (95). Because empathy is considered a ‘feminine’ trait based on the response from the survey, it is considered inadequate for a high leadership position rather than being perceived as a strength.

Conversely, for weaknesses for female leaders, one respondent stated, “However, sometimes they could be a bit too harsh on themselves in terms of trying to make things as perfect as possible so there are a lot of adjustments to reach the end product.” This shows that their overachieving and perfectionist mindset is what hinders them from taking on big roles — they fear that they will be disappointed in the future, and they are scared of the potential failure. Another weakness is the social bias against females, as a response reads, “My personal experience working with female leaders was also pretty good, but I can't really see many big organizations with females as leaders.” This portrays how the stereotypical views of society don’t normalize females as leaders. This impedes the female leaders when opportunities are offered to them, as they fear that they will not be accepted by the men-dominated society. This depicts the reality of the stereotypical society, where females are not perceived to be adequate for

top-level leadership positions despite their acknowledged competence. The last weakness is their lack of confidence. One person asserted, "Female leaders may struggle with assertiveness and confidence and may take longer to make crucial decisions." This shows that women's position in our society might not concern their competence at all, and that their lack of self-assurance and confidence is detrimental to their situation.

Synthesis of Results with the Literature

The Confidence Gap

Throughout this study, the results point to the confidence gap as a factor for the paradox between the underrepresentation of female leaders and the perception of female leadership. This is displayed in the study when women were ranked higher in leadership, but they were stated to lack the traits of 'confidence' or 'risk-taking.' The confidence gap refers to a difference in self-assurance between men and women, especially when it comes to professional settings. Past research has shown that despite equal performance, women rate their skill and potential as lower than men, which affects how readily they take on challenges and seek advancement in their firms (Zenger). The confidence gap is not about women's real capabilities in their roles, but the self-assurance they have in their abilities.

In addition, Hewlett-Packard (HP) found that men will apply for a job or promotion when they meet only 60% of the qualifications, but women will only apply if they meet 100% of them (HP Inc.). This aligns directly with the results from the survey results, where 85.2% of respondents linked risk-taking to men, which shows that females are less likely to take the risk that they need to get promoted and advance in their positions. Ultimately, this research highlights that a paradox exists where students recognize the reliable competence of female leaders, but they reserve the traits of confidence and risk-taking for men, which are the traditional characteristics that align with executive workforce promotions.

The Imposter Syndrome

The survey did not directly mention imposter syndrome, but it can be inferred that the underrepresentation of women can be attributed to imposter syndrome. Imposter syndrome refers to a case where highly competent individuals tend to doubt their abilities and accomplishments in fear of not belonging to a particular environment or position. Past research has suggested that women experience impostor phenomenon more frequently and intensely than men do (Price et al.), which shows that women are more likely to hesitate from progressing to higher positions even when they are capable of it.

This raises an important question: Why does this phenomenon occur at much higher rates for female workers? Research shows that when a female leader fails in a high position, they are much more likely to be judged more severely due to the bias of society (Elsesser). Because the penalty for failure is so much higher, women naturally feel they must be completely certain

whether they can bring success to their role before stepping up for a position. This links to another idea of how gender biases can influence the role of females in the workforce.

Because the survey measures how students perceive leaders rather than how they perceive themselves, it does not observe the imposter syndrome directly. However, the imposter syndrome is reflected in the results to a certain extent, because 85.2% of respondents voted for men when it came to 'Innovational/Risk Taking.' This supports the belief that a major barrier to female leadership is their self-doubt, compared to males who are less likely to suffer from imposter syndrome.

Transactional and Transformational Leadership

This study is associated with the notion of two distinct leadership styles, transformational and transactional leadership. Transactional leadership refers to a task-based leadership style that uses structure, compliance, and rewards. Transformational leadership style aims to inspire the subordinates by sharing a long-term vision, innovation, and motivation. The students' responses in the study demonstrate the association of males with transformational leadership, as the majority of the respondents selected 'Male' when identifying the efficacy for 'Motivating others with Inspiration' and 'Innovational/Risk Taking'. On the other hand, females were correlated with transactional leadership by the students, since they were selected for the categories 'Supervision and Accountability', 'Logistics and Planning', and 'Rewards Management Style.'

These results contradict past studies. According to a study by Feroz Khan in 2024, women are more likely to adopt a transformational leadership style driven by inspiration, motivation, and creating relationships with team members. Contrastingly, men are likely to adopt a transactional style that associates with structure, task, rewards, and goals (F. Khan). This contradiction suggests how the high school students are judging the leadership traits on persistent stereotypes and patterns in the corporate world. They are attributing the transformational innovation and vision to men, in accordance with the high percentage of male global CEOs. This further highlights how the cognitive bias is formed from early stages of development, setting a barrier for women when taking on executive roles.

Gender Bias and Stereotype

Gender bias and stereotypes also have a big influence on how females gain access to opportunities in the workforce. One response observes, "I think that in a society where misogynistic attitudes still exist, the unfortunate reality is that it's harder for women to lead large groups of people, including men." This respondent highlights a systemic and cognitive barrier in society. This respondent is not only identifying the bias, but also identifying the pattern in others' attitudes and perspectives that hinders female leaders' opportunities. This aligns with existing research because one study stated, "Women who hold leadership positions offer distinctive viewpoints that contribute to workplace positivity, innovation, and decision-making. Nevertheless, they encounter significant obstacles, such as underrepresentation, bias, and restricted opportunities for advancement" (Setyaningrum and Juansih 252). Therefore, the

integration of my study with past research shows that this underrepresentation and bias towards women is not only present in contemporary corporate leadership but also stems from the perceptions of the young minds.

Conclusion

The results of this study highlight a paradoxical relationship between the proven competence of the female leaders and their representation in leadership positions. To address the initial research question of this study, 'To what extent do the perceptions of Vietnamese high school students on gender and leadership align with the gender imbalance in corporate leadership roles?', the results of the questionnaire revealed that the perceptions of the high school students in Vietnam do not align with corporate leadership metrics, highlighting the underrepresentation of female leaders. The results also allow for an area of further understanding on how the problem is caused by factors such as the confidence gap and imposter syndrome. To fully address this persistent inequality, future researchers should aim to identify whether this paradox is persistent across young students in other nations as well, expanding beyond the scope of Vietnam. Ultimately, until the cognitive biases on different genders shift and women gain more confidence in their capabilities, the underrepresentation of females in leadership will persist.

Works Cited

Abrams, Jennifer. "Women in Leadership: Gender Bias and the 'Confidence Gap.'" Voice Lessons, LLC, May 5, 2015, jenniferabrams.com/women-in-leadership-gender-bias-and-the-confidence-gap/. Accessed June 28, 2026.

Catalyst. "Women CEOs." Catalyst, 2022, www.catalyst.org/insights/featured/women-ceos. Accessed May 2, 2026.

Eagly, Alice H., and Steven J. Karau. "Role Congruity Theory of Prejudice toward Female Leaders." Psychological Review, vol. 109, no. 3, 2002, pp. 573–98. https://doi.org/10.1037/0033-295X.109.3.573.

Elsesser, Kim. "Gender Bias against Female Leaders: A Review." Handbook on Well-Being of Working Women, edited by Mary L. Connerley and Jiyun Wu, Springer, 2016, pp. 161–73. https://doi.org/10.1007/978-94-017-9897-6_10.

Field, E., Krivkovich, A., Kügele, S., Robinson, N., & Yee, L. (2023). Women in the Workplace 2023. In McKinsey & Company. McKinsey & Company. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/women-in-the-workplace

Freeman, R. Women-Owned Enterprises in Vietnam: Perceptions and Potential. IFC, 2017, www.ifc.org/content/dam/ifc/doc/mgrt/market-study-on-women-owned-enterprises-in-vietnam-eng-v1.pdf.

Grant Thornton. Women in Business 2025. Grant Thornton Vietnam, 2025, www.grantthornton.com.vn/insights/thought-leadership/women-in-business/women-in-business-2025-for-Vietnam-market/.

Hoobler, J. M., et al. "Women's Managerial Aspirations." Journal of Management, vol. 40, no. 3, 2014, pp. 703–30. https://doi.org/10.1177/0149206311426911.

HP Inc. HP Global Survey on Workplace Trends / Gender Equity. Morning Consult, 2022, www.hp.com/content/dam/sites/garage-press/press/press-blogs/2022/gender-equity-workplace-study/Global%20-%20HP%20IWD%20Data%20Fact%20Sheet%20-%20Final.pdf.

Khan, Feroz. "Gender Differences in Leadership Styles and Organizational Success in Vietnam." European Journal of Gender Studies, vol. 6, no. 2, 2024, pp. 11–21. ideas.repec.org/a/bfy/ojejgs/v6y2024i2p11-21id2238.html.

Khan, Walayet A., and João Paulo Vieito. "CEO Gender and Firm Performance." Journal of Economics and Business, vol. 67, 2013, pp. 55–66. https://doi.org/10.1016/j.jeconbus.2013.01.003.

Kruse, Kevin. "New Research: Women More Effective than Men in All Leadership Measures." Forbes, March 31, 2023, www.forbes.com/sites/kevinkruse/2023/03/31/new-research-women-more-effective-than-men-in-all-leadership-measures/.

Lien, T. T. H., and P. T. T. Thuy. "Impacts of Gender Diversity on Corporate Performance: A Study of Board of Directors and Top Management." Corporate Board: Role, Duties and Composition, vol. 20, no. 2, 2024, pp. 89–101. https://doi.org/10.22495/cbv20i2art9.

McKinsey & Company. Women in the Workplace. 2024, www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace.

Meynell, L. "More Women Take CEO Jobs but Parity Still Decades Away." Russell Reynolds Associates, April 13, 2023, www.russellreynolds.com/en/insights/articles/ceo-turnover-insights. Accessed May 2, 2026.

Price, P. C., et al. "Gender Differences in Impostor Phenomenon: A Meta-Analytic Review." Current Research in Behavioral Sciences, vol. 7, 2024, article 100155. https://doi.org/10.1016/j.crbeha.2024.100155.

Schein, Virginia E. "The Relationship between Sex Role Stereotypes and Requisite Management Characteristics." Journal of Applied Psychology, vol. 57, no. 2, 1973, pp. 95–100.

Setyaningrum, Ganis, and Juansih. "Gender-Based Challenges in Women's Leadership Careers: A Literature Synthesis." Technium Social Sciences Journal, vol. 63, 2024, pp. 252–67. https://doi.org/10.47577/tssj.v63i1.11929.

Tran, T. T. T., and H. V. Nguyen. "Gender Preference in Higher Education Leadership: Insights from Gender Distribution and Subordinate Perceptions and Expectations in Vietnam Universities." International Journal of Leadership in Education, 2020, pp. 1–22. https://doi.org/10.1080/13603124.2020.1753244.

Zenger, Jack. "The Confidence Gap in Men and Women: How to Overcome It." Zenger Folkman, Jan 25, 2021, zengerfolkman.com/articles/the-confidence-gap-in-men-and-women-how-to-overcome-i

Zenger, Jack, and Joseph Folkman. "Research: Women Score Higher than Men in Most Leadership Skills." Harvard Business Review, June 25, 2019, hbr.org/2019/06/research-women-score-higher-than-men-in-most-leadership-skills. Accessed May 2, 2026.

A Comparative Modal Analysis of Automotive Suspension Systems By Yueqian Jiang

Abstract

Suspension plays a vital role in vehicle comfort, handling, and safety, yet traditional and modern suspension systems are often discussed separately despite being governed by the same underlying principles. This paper presents a pedagogical approach for understanding automotive suspension dynamics through the concept of natural frequency. Beginning with a single degree of freedom (SDOF) mass–spring–damper system the analysis introduces the basic principles of mass, stiffness, and energy dissipation, and how they collectively influence vehicle behavior. These principles are then applied to explain the most common vehicle suspensions using the MacPherson strut as a representative example. Building on this foundation, three types of modern suspension systems (magnetorheological suspension, air suspension, hydraulic suspension) are examined for how they dynamically modify stiffness and damping to alter vehicle response in real time. Despite their increased complexity, these systems can be understood using the same foundational vibration concepts, providing a consistent framework for analyzing a wide range of vehicle suspension technologies.

Viewed through a modal framework applied to the MacPherson strut, these systems form a clear functional taxonomy. Magnetorheological systems primarily vary damping while maintaining the designed natural frequency to preserve handling balance and suppress resonant peaks. Air suspension primarily varies stiffness, directly shifting natural frequency to adapt ride characteristics under changing loads and operating conditions. Hydraulic systems can vary both stiffness and damping while exerting net forces, enabling control over modal behavior, eigenvalues, and coupling. Different strategies trade adaptability and performance for increased system complexity.

Introduction

The ride comfort, handling stability, and safety of a vehicle are determined by the responses of its suspension system to road-induced disturbances. Road irregularities generate vibrations across a wide range of frequencies that are transmitted through the suspension system. The prediction and control of such responses are critical issues for vehicle engineering. Vibration theory, as applied to vehicle systems, provides a framework that links vehicle performance measures to system parameters. Excessive vibration can compromise driver control, increase fatigue, and elevate the risk of accidents, directly impacting passenger safety. In addition, poor vibration management can accelerate component wear, leading to higher maintenance costs and reduced vehicle lifespan, which has clear economic consequences for both manufacturers and consumers. This economic significance is further reflected in the scale of the industry, with the global automotive suspension system market valued at approximately 151.71 billion USD in 2026 and projected to reach 204.46 billion USD within five years (Mordor Intelligence).

Natural frequency is a fundamental property of a vibrating system that represents its inherent frequency of oscillation. When an external input excites a system near its natural

frequency, large responses can occur, which may lead to excessive vibration or even structural damage. Therefore, it is important to determine and control the natural frequency of a vibration system. In vehicle dynamics, natural frequency couples vibration theory directly to suspension system behavior.

Suspension systems are commonly modeled using a lumped-parameter approach, which is useful for modeling. Early studies of vehicle suspensions typically considered simple models, such as single-degree-of-freedom (SDOF) and two-degree-of-freedom (2-DOF) models, to investigate various aspects of vehicle suspensions, such as mass, stiffness, and damping, under harmonic excitations (Meirovitch; Rao and Griffin). These studies established how the underlying concepts—that undamped natural frequency is primarily governed by the stiffness and mass ratio, and that damping regulates the amplitude of vibration rather than frequency—are applied to automotive suspension design. Later, several studies extended this approach and proposed new ideas, such as resonance, transmissibility, and frequency response functions, to quantify vehicle vibration isolation and force transmissibility (Thomson; Blevins).

For vehicle suspension, the initial studies have shown that driver perception of vehicle dynamics has a close relationship to the sprung mass natural frequency and damping ratio (Karnopp; Gillespie). A lower sprung-mass natural frequency improves isolation from road disturbances but reduces handling responsiveness, whereas a higher natural frequency enhances control at the expense of ride comfort. This fundamental trade-off governs suspension tuning (Rajamani). More recent studies have also shown that the ability to mitigate the effects of vibration depends on the control of system transmissibility (Wong).

Quarter car models have further separated the dynamics of the sprung and unsprung masses. Using them researchers have shown that the presence of wheel-tire stiffness introduces a higher frequency mode and that the magnitude of the unsprung mass affects vibration transmission (Goodarzi and Khajepour). These studies have also shown that commonly used suspension geometries have a direct impact on the effective wheel rate and therefore the system's natural frequency (Bucchi and Lenzo).

Advances in sensors, control systems, and actuators have led to the development of suspension systems with variable dynamics. Semi-active and active suspensions have improved the system's ability to control vehicle behavior. For example, magnetorheological dampers regulate the damping coefficient, which affects vibration amplitude and system response (Soliman and Kaldas; Wang et al.). Air suspension systems allow adjustment of stiffness, enabling the system's natural frequency to shift under different operating conditions (Moheyeldein et al.). Active hydraulic suspension systems further extend this capability by reshaping the system’s eigenvalues through simultaneous control of stiffness and damping matrices (Guo et al.; Ling et al.), thereby emphasizing the continued relevance of classical vibration theory.

Although these technologies are often discussed independently focusing on specific areas such as modeling, implementation, or control, they can be understood within a common modal framework based on natural frequency theory. Viewing suspension systems through this

framework clarifies how mass, stiffness, damping, and geometry govern dynamic response across passive, semi-active, and active architectures.

To develop this unified framework, this paper begins with a discussion of a single degree-of-freedom mass-spring-damper system, which establishes the core concepts of undamped and damped natural frequency, resonance, damping ratio, transmissibility, and quality factor. These concepts are then applied to explain how a strut-based suspension system functions in practice. With this mechanical foundation established, more modern semi-active and active suspension technologies are examined for how they modify system behavior and response characteristics within the same modal framework.

Background

Suspension dynamics can first be understood using a single-degree-of-freedom (SDOF) mass-spring-damper system to define natural frequency with and without damping, resonance frequency, damping ratio, transmissibility, and the quality factor, while examining how mass, stiffness, and damping influence the system's frequency response and overall suspension performance.

Natural frequency refers to the rate at which a system oscillates when disturbed from equilibrium without any external forces. In the context of automobile suspension systems, natural frequency governs the response of the system to road irregularities, which affects both ride comfort and handling stability of the vehicle (Karnopp 45–75). For a simple undamped system, the natural frequency is determined by the mass and stiffness. The undamped natural frequency ω_n is

$$\omega_n = \sqrt{\frac{k}{m}}, \tag{1}$$

where k is the spring stiffness (N/m) and m is the mass (kg) of the suspended body. Increasing stiffness or decreasing mass raises the natural frequency. This causes the suspension to respond more quickly, but it also transmits more vibration to the passengers (Juvinall and Marshek 110–35).

When external forces act periodically, the system's motion can be modeled by:

$$F(t) = F_0 \cos cos\ (\omega t). \tag{2}$$

In equation 2, $F(t)$ is the applied force at time t, F_0 is the amplitude of the force, and ω is the forcing frequency (rad/s). In suspension terms, such forcing may arise from repeated wheel interactions with road irregularities, such as bumps or undulations (Meirovitch 85–120).

Resonance occurs when the forcing frequency equals the natural frequency ($\omega \approx \omega_n$). This condition produces large oscillations that degrade ride quality and vehicle control. Suspension designers therefore set ω_n away from the major excitation frequencies encountered in typical driving (Nayfeh and Mook 15–35).

A single-degree-of-freedom (SDOF) system with damping is modeled with damping, represented by the damping ratio ζ, so that the equation of motion of an SDOF damped system is:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F_0 \cos cos\,(\omega t). \quad (3)$$

In equation 3, x is the displacement, $\dot{x}$ is the velocity, and $\ddot{x}$ is the acceleration; c is the damping coefficient (N·s/m). Damping dissipates vibratory energy and suppresses oscillation amplitude, particularly near resonance (Rao and Griffin 140–85).

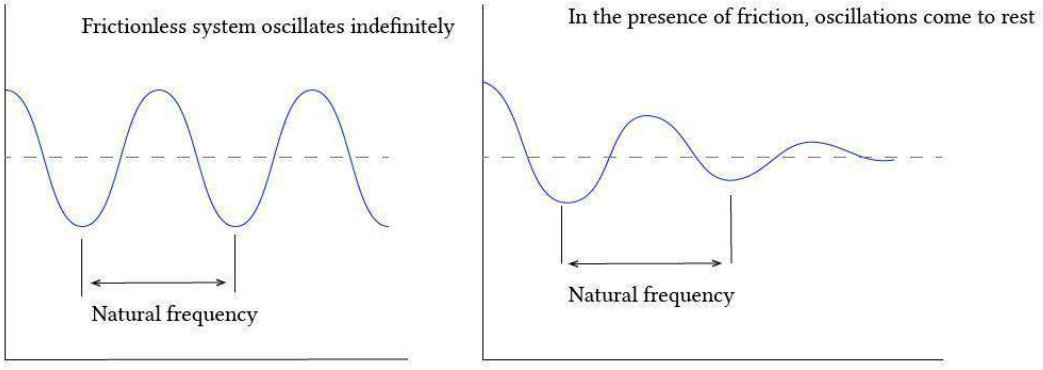


Fig 1: Illustrative Diagram of Damped vs Undamped Oscillations; The X-axis represents time while the Y-axis represents position, with the dashed line showing the center of the motion. Reprinted from Alex Yartsev, *Resonance, Damping and Frequency Response*, Deranged Physiology 2024.

The damped natural frequency ω_d is

$$\omega_d = \omega_n\sqrt{1 - \zeta^2}. \quad (4)$$

Damping reduces the oscillation frequency below the undamped value. This property is critical in suspension tuning: it prevents the system from resonating with road inputs (Wong 245–75).

In practice, vibrations caused by road surfaces typically contain a range of frequencies. The transmissibility coefficient (Tr) is measured as a percentage of input vibration transmitted via suspension to the body of the vehicle:

$$T_r = \frac{\sqrt{(2\zeta r)^2 + \left(1 - r^2\right)^2}}{\sqrt{\left(1 - r^2\right)^2 + (2\zeta r)^2}}, \quad (5)$$

where $r = \frac{\omega}{\omega_n}$ is the ratio frequency. Suspension systems are tuned so that T_r decreases within the frequency range most critical to ride comfort (Thomson 190–225).

Finally, the quality factor (Q) measures how underdamped the system is: $Q = \frac{1}{2\zeta}$. Higher Q produces higher peaks of resonance, which is undesirable in suspension for the sake of comfort but might be desirable for performance use where fast response is desirable (Blevins 1–25).

Taken together, natural frequency, damping ratio, and transmissibility collectively govern suspension system dynamic behavior. Accurate tuning ensures the system will effectively dampen road unevenness without producing excessive vibrations, achieving a compromise between comfort, safety, and handling capability.

In summary, natural frequency, damping ratio, transmissibility, and the quality factor collectively describe how a suspension system responds to road inputs across a range of frequencies. Understanding these relationships within the SDOF framework is essential before examining how they scale and interact in more complex vehicle suspension architectures.

Traditional Suspension: The MacPherson Strut

The SDOF principles established in the preceding section now scale directly into real vehicle design. The following discussion applies natural frequency and damping concepts to the MacPherson strut suspension system. The spring, damper, tire, and suspension links can be modeled as parts of an interacting mass-spring-damper system. In vehicle dynamics, this behavior is commonly represented using quarter-car and multi-degree-of-freedom models, which help describe how sprung mass, unsprung mass, stiffness, damping, and tire compliance influence ride and handling. The fundamental equations of damping and natural frequency are applied to vehicle suspension design, with emphasis on ride comfort and handling performance.The MacPherson strut, a common front suspension architecture, is used as the main example to show how motion ratio, wheel rate, unsprung mass, tire stiffness, and damper tuning affect the vehicle’s dynamic response.

A contemporary passenger vehicle suspension system consists of four linked functional subsystems used to manage ride comfort and dynamic rollover stability at higher speeds: elastic elements (such as springs), energy-dissipation elements (such as shock absorbers or dampers), the wheel-tire assembly, and kinematic links (such as control arms, uprights or knuckles, struts or wishbones, anti-roll bars). Together, these subsystems represent a coupled mass-spring-damper system, in which the sprung mass (the vehicle body supported by the suspension's elastic members) and the unsprung mass (including the wheel, tire, brake system, and suspension elements not supported by the spring) interact through spring stiffness, damping properties, and geometrically defined motion ratios.

In vehicle dynamics, this behavior is commonly represented using simplified models such as the quarter-car and two-degree-of-freedom (2-DOF) systems. A quarter-car model idealizes one wheel, suspension, and body portion, which is equivalent to one-fourth of the vehicle's mass, commonly used for studies on vertical vibrations. Within the two-degree-of-freedom model, vertical movements of the unsprung and sprung masses can be varied independently, which helps to predict different body and wheel modal or natural frequencies and the corresponding transfer of road vibration inputs to body acceleration and tire deflection, respectively (Rajamani 160–220; Gillespie 90–140).

In suspension design, the spring governs the static load and largely determines the natural frequency of the supported mass; dampers regulate transient energy dissipation and suppress

damp peak resonant response; tires add a relatively stiff component influencing higher frequency wheel modes; and linkage kinematics define effective wheel stiffness (wheel rate) through the motion (installation) ratio (Goodarzi and Khajepour). Together, these physical functions provide the basis for quantitative interrelationships that will be discussed.

The undamped natural frequency of a single linear mass–spring oscillator is the primary relation used when sizing and tuning automotive suspensions. For a SDOF the angular natural frequency is $\omega_n = \sqrt{\frac{k}{m}}$, and the corresponding frequency in cycles per second is $f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$. When damping is present, the damping ratio is defined by $\zeta = \frac{c}{2m\omega_n}$ and the damped oscillation frequency is $f_d = f_n\sqrt{1 - \zeta^2}$. The coefficient for critical damping may be written as $c_c = 2m\omega_n = 2\sqrt{km}$. These relations form the foundation for selecting spring rates and damper characteristics in vehicle engineering and are used directly in quarter-car models of ride and isolation (Rajamani 160–220; Goodarzi and Khajepour).

Suspension designers translate ride comfort into defined modal characteristics, most notably the undamped sprung-mass natural frequency, often referred to as ride frequency. In passenger cars ride frequency is generally targeted at around 1.0 Hz where a balance between comfort and control is achieved. Larger trucks and heavy-duty vehicles require higher resonant frequencies to satisfy their load-carrying requirements (White). A reduced target ride frequency (approximately within the range of 0.8 Hz to 1.2 Hz) increases isolation of road irregularities but lowers roll and pitch stiffness, whereas a higher ride frequency (greater than 1.5 Hz) increases responsiveness and agility but at the cost of a harsher ride feel (Rajamani 160–220). The selected ride frequency therefore determines the required wheel rate and, through suspension geometry and motion ratio, the corresponding spring rate used in the system.

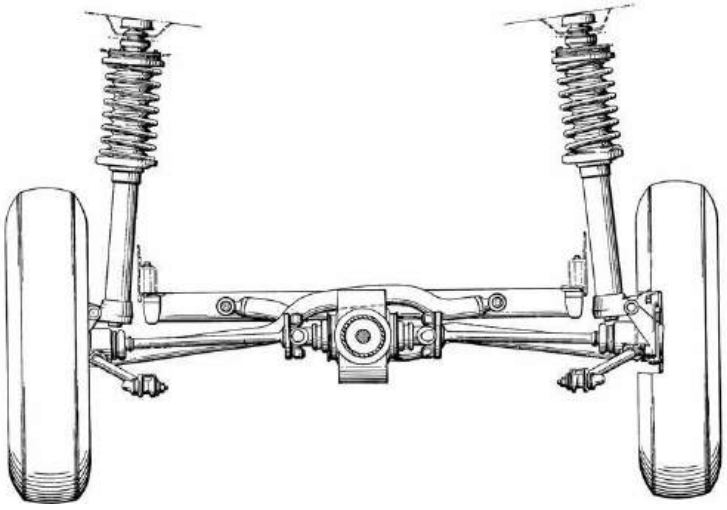


Fig 2: Diagram of the MacPherson Strut; This illustration shows the components and relative positions of a standard MacPherson strut. Adapted from E. S. MacPherson, *Vehicle Wheel Suspension System*, 1953, U.S. Patent No. US2624592A

The MacPherson strut (shown above)is a widely used suspension architecture that integrates a coil spring coaxially around a telescopic damper into a single vertical assembly, combined with a lower control arm and steering knuckle. Its widespread use results from its compact packaging, reduced component count, and efficient load path. The kinematic behavior of the strut—particularly camber variation, roll-center location, and motion ratio—governs the effective wheel rate and therefore the system's natural frequency during operation.

The suspension at a single wheel is commonly modeled as two-degree-of-freedom quarter-car model: a sprung mass m_s connected to the wheel through an effective wheel spring constant K_w and a damper C_w, and the wheel and tire unit represented as an unsprung mass m_u and a tire stiffness K_t. Since the MacPherson spring and damper are mounted directly upright, the motion ratio (r_{MR}) normally tends to unity; however, since it is a function of travel and steering angle owing to upright geometry, it varies. This variation introduces mild nonlinearity in the wheel rate across the working range (Bucchi and Lenzo). The effective wheel rate is linked to the strut spring constant through the motion ratio, given by $K_w = K_s \cdot \left(r_{MR}\right)^2$, hence allowing the designer to determine the strut coil rate K_s by inverting this relationship after determining the desired wheel rate based on the aimed sprung natural frequency.

Motion ratio and wheel rate determine how spring stiffness couples into the dynamic response experienced by occupants. Because motion ratio depends on suspension geometry and varies with travel, the effective wheel rate and therefore the ride frequency exhibits mild nonlinear behavior during operation.

Distribution and quantity of unsprung mass in MacPherson configurations are also typically larger than some multi-link designs since the strut concentrates additional assembly mass near the wheel. Increased unsprung mass raises the unsprung-mass mode frequency and transmits higher-frequency accelerations directly to the body during tire-excitation events.

Roll-center movement and camber change shift the load distribution during transient maneuvers, modulating the apparent sprung response under combined excitations. Hardpoint location is therefore used to tune both the handling response and the apparent ride frequency during dynamic events. Damper strategy, which is the damper generating force as a product of both velocity and displacement, governs the damping ratio ζ of the sprung mode, which in turn dictates resonant-peak magnitude and transient settling time. On passenger cars, moderate values $\zeta \approx 0.2 - 0.4$ are typical for the bounce mode, indicating the compromise between comfort and control (Rajamani 160–220; White; Bucchi and Lenzo; Tapia et al.).

In practice, the design process starts off choosing a target sprung natural frequency $f_{n,s}$ (for example, approximately 1.0 Hz for a standard passenger car) and calculates the required wheel rate for the corner from the relationship

$$f_{n,s} = \frac{1}{2\pi}\sqrt{\frac{K_w}{m_{corner}}}. \quad (6)$$

Given an assumed corner sprung mass m_{corner}, equation 6 yields K_w. The physical spring rate at the strut is then chosen so that $K_w = K_s \cdot \left(r_{MR}\right)^2$, based on the motion ratio obtained during kinematic analysis. Damping is chosen to provide the necessary damping ratio $\zeta = \frac{c}{2m\omega_n}$, where c is the damper coefficient, ω_n being the angular natural frequency. Since the MacPherson integrates the spring axis, damper axis, and steering knuckle into a single assembly, the design procedure must combine kinematic analysis (to provide motion ratio behavior, together with the roll/camber behavior, over travel) with classical modal calculations. Comparative studies confirm that installation-ratio expressions and kinematic layout directly govern wheel rate and the resulting natural frequencies, particularly when comparing MacPherson to multi-arm suspensions (Bucchi and Lenzo; Tapia et al.).

The SDOF principles from the Background section translate directly into engineering practice for the MacPherson strut. Under the relationships $\omega_n = \sqrt{\frac{k}{m}}$ and $f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$, along with wheel-rate conversion $K_w = K_s \cdot (MR)^2$designers convert ride objectives (approximately 1 Hz) into tangible spring and damper specifications as well as geometric hardpoint choices that determine motion ratio as well as camber behavior. The MacPherson strut stands the test of time due to packaging as well as structural integrity that keep these compromises within manageable limits; nevertheless, kinematic behavior as well as the implications of unsprung mass need to be accounted for explicitly during natural frequency tuning to optimize ride comfort as well as dynamic stability (Rajamani 160–220; Bucchi and Lenzo; Tapia et al.).

In summary, this section shows how the essential mass-spring-damper equations govern practical design decisions—ride frequency targets, wheel rate, motion ratio, and damping strategy—as unified in the MacPherson strut. These same parameters become the variables that modern active and semi-active systems manipulate in real time, which the following section examines.

Modern Advancements in Suspension

Having established the fixed-parameter behavior of passive MacPherson-type suspension systems, this section examines how modern active and semi-active systems manipulate those same parameters in real time to improve comfort, roadholding, and safety. These systems shift natural frequency and regulate modal response by dynamically modifying key suspension parameters, specifically stiffness (k) and damping (c)—which in turn alters the system's natural frequency and modal response ω_n and ζ. Three different types of systems will be examined, including magnetorheological dampers, air suspension, and hydraulic active suspension.

This section applies the natural-frequency framework developed in the preceding sections to compare these three active and semi-active technologies. In passive systems such as the MacPherson the corner possesses an essentially fixed effective stiffness, K_w, such that the

predictable sprung natural frequency, $\omega_n = \sqrt{\frac{K_w}{m}}$, ensues, where the damping, $\zeta = \frac{c}{2m\omega_n}$, cannot be varied other than if the damper is variable. Each technology examined below modifies one or both of these parameters: magnetorheological dampers vary c; air springs vary k; hydraulic systems can vary both k and c simultaneously.

Magnetorheological (MR) Semi-Active Suspensions

Magnetorheological (MR) semi-active suspensions center on an MR fluid-based damper whose apparent viscosity changes almost instantaneously as the magnetic field is applied across the piston. The system consists of a piston or valve set embedded coil (electromagnet), the MR fluid, a high-rate current driver as well as an electronic controller that uses vehicle sensors (accelerometer, wheel sensors, steering, speed) to change current on the fly.

When a magnetic field is applied, ferrous particles within the fluid align along the field lines, increasing the fluid's yield stress and the damper force versus piston velocity. This behavior raises the effective damping coefficient c directly while leaving the mechanical spring stiffness k nominally constant (Wang et al.).

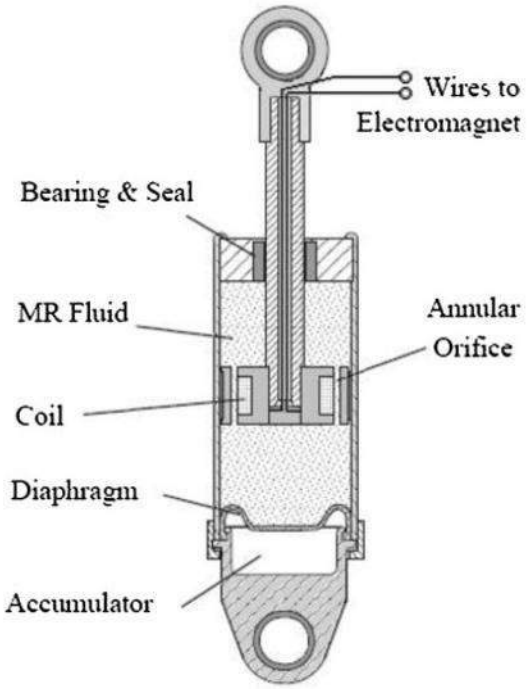


Fig 3: Diagram of a Magnetorheological (MR) Suspension; This illustration shows the components, their names, and relative locations in an idealized MR suspension. Reprinted from I. Bica et al., *Physical Characteristics of Magnetorheological Suspensions and Their Applications*, Journal of Industrial and Engineering Chemistry 2013.

Because the MR strategy modifies only damping, it varies the damping ratio without substantially shifting the angular natural frequency ω_n. Natural frequency depends on k and m

according to $\omega_n = \sqrt{\frac{k}{m}}$, so it remains stable. Raising c increases $\zeta = \frac{c}{2m\omega_n}$, which minimizes resonant amplification within the frequency response function (FRF) and reduces transient ringing without shifting the resonance frequency location. This is the reason why semi-active MR devices are so prevalent as damping regulators: they suppress peak responses in the frequency response function while retaining the baseline corner stiffness, handling balance, and motion ratio (Soliman and Kaldas).

Magnetorheological damping is widely used in production performance vehicles, where it is employed to enhance both ride comfort and handling. In these applications, wheel rate K_w establishes the natural frequency ω_n while the MR damper actively varies the damping coefficient c to control the amplitude of the resonant response as well as of the transient settling time without substantially altering ω_n (Wang et al.). This allows the system to reduce vibration under road disturbances while preserving the handling characteristics governed by the suspension geometry. From a comfort perspective, MR systems reduce peak body accelerations at ω_n which are experienced as more jarring. From a handling perspective the technology holds the baseline corner eigenvalue which helps to preserve the desired ride-response balance intended. From a safety perspective, the high rates of MR dampers enable control authorities to rapidly raise damping when inputs are vigorous, restoring vehicle stability. Disadvantages are that cost is higher and electronic complexity is higher compared to passive dampers. MR systems also cannot change ω_n when a different ride frequency would be beneficial, because k is constant (Soliman and Kaldas).

Air Suspension

Air suspension systems replace or supplement traditional coil springs with pressurized bellows supported by a compressor, valves, reservoirs, and a control unit charged with controlling pressure and effective volume. The primary effect is that adjusting pressure directly governs the effective spring stiffness k. In some multi-chamber types, nonlinear stiffness properties are adjustable through changes in effective volume. This configuration enables air suspension mechanisms to have a large range of k variability such that changes to stiffness component operation can be made to ω_n (Moheyeldein et al.).

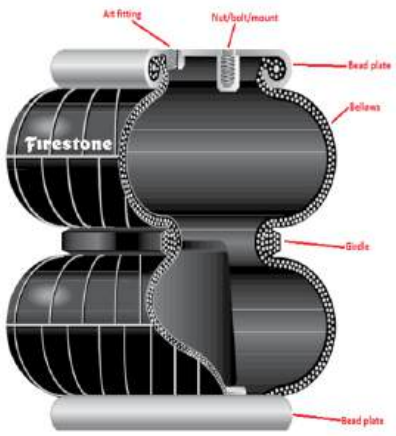


Fig 4: Diagram of an Air Suspension; This illustration shows a cutaway of a standard air suspension, its components, and their names. Reprinted from GlobalSpec, *Air Springs Information*, 2026.

Because natural frequency varies with the square root of stiffness, changes in k directly shift the system's dynamic response. Increasing stiffness by a factor α shifts the new angular frequency to $\omega_n^{'} = \sqrt{\frac{\alpha k}{m}} = \sqrt{\alpha}\,\omega_n$. Modern air systems on high end cars employ closed loop control and occasionally multi chamber bellows to keep passenger experienced ride frequency close to some comfort target as load varies and to increase ω n for sport modes when higher stiffness is better handling (Moheyeldein et al.).

Operationally, air suspension acts as a direct lever on the eigenvalue locations in the vehicle's stiffness matrix. Increasing pressure raises k and shifts ω_n up which is desirable when body motion needs to be minimized to improve handling. Conversely, decreasing pressure lowers k and shifts ω_n downward, which improves isolation from road roughness. Because air systems modify a stiffness entry in the vehicle's stiffness matrix, they can shift modal eigenvalues to suit payload or driver preference in ways that damping adjustment alone cannot achieve (Moheyeldein et al.).

From a comfort perspective, air suspension systems are most efficient when maintaining a low perceived ride frequency against changing payloads and consequently enhance passenger comfort and self-levelling. They also offer the ability to retarget ω_n to achieve a sportier character, which can enhance turn-in and body control. Safety advantages are ride height maintenance and predictable body attitude in an emergency. Disadvantages are higher mechanical complexity from valves and compressors, slower large amplitude response than fast MR damping changes, and susceptibility to air leaks which requires strong sealing and control techniques. Selective integration with damping and vehicle control functions is also necessary to prevent transient stiffness-damping mismatches that could temporarily unsettle handling (Moheyeldein et al.).

Hydraulic Active Suspension

Hydraulic active suspension systems range from semi active valve control to complete active architectures that exert actuator forces through hydraulic cylinders, pumps, and electronic valves. Primary hardware consists of high-pressure pumps, hydraulic actuators, energy buffering accumulators and a central controller as well as fast valves. Their distinguishing capability is near simultaneous, high-bandwidth control of both of effective stiffness k—through interconnections and active force gradients—and damping c through valve modulation. This enables direct manipulation of the system's dynamic response, extending beyond passive and semi-active limitations. Fully active systems have the potential to exert net forces larger than passive k and c behavior, and can therefore eliminate or shift modes rather than simply suppress them (Guo et al.).

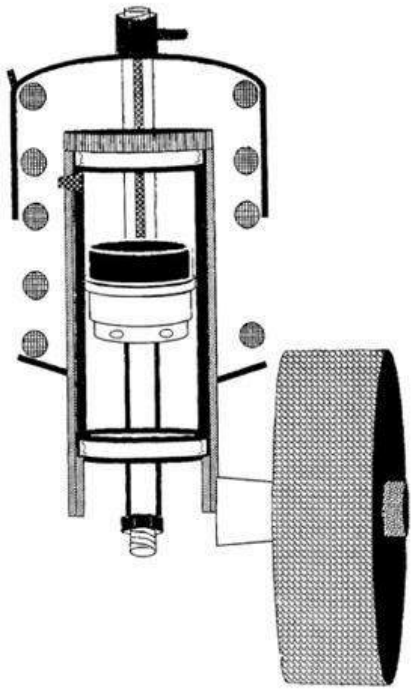


Fig 5: Diagram of Hydraulic Suspension; This illustration shows the standard components of a hydraulic suspension. Adapted from P. D. Maria, *Active Suspension System and Hydraulic Ram Therefore*, 2013, U.S. Patent No. US8573605B2.

Because hydraulic architectures can modify entries in both the stiffness matrix and the damping matrix, they independently shift eigenvalues (changing ω_n) and modal response (changing ζ). When system interconnectivity is employed, topological coupling across corners enables active modification or suppression of global body modes—heave, pitch, and roll. Mathematically, this action results in changing the effective K and C matrices and shifts the eigenvalues $\omega_{n,i}$ and of damping ratios ζ_i of the overall vehicle system (Ling et al.).

Manufacturers employing hydraulic actuation target applications requiring both high levels of ride comfort and precise vehicle control. Hydraulic actuators can apply forces that

cancel body motions, driving perceived ride frequency to very low values on poor roads. Conversely, the same hardware can produce high virtual stiffness for cornering, raising ω_n and improving handling. Because hydraulics can govern both stiffness and damping simultaneously, they offer the broadest modal control of the three families considered here (Guo et al.; Ling et al.).

From a comfort perspective, hydraulics can achieve very low passenger perceived motion through active cancellation of body modes. Handling benefits follow from the ability to stiffen the vehicle dynamically, redistributing load during cornering. Safety benefits include active control of roll and pitch, which increases tire contact during extreme maneuvers. Drawbacks include complexity, weight, power requirement, cost, maintenance burden, and potential single-point failures such as loss of hydraulic pressure. These risks require redundancy and fail-safe protection. Additionally, careful control design is required because inappropriate actuator commands can excite undesirable modes when vehicle dynamics modeling and sensor feedback are insufficient (Guo et al.; Ling et al.).

Conclusion

Viewed through the modal framework developed across this paper, the three technologies constitute a clear taxonomy rooted in how each modifies the fundamental parameters of the mass-spring-damper model. Magnetorheological damping primarily varies c and ζ while holding ω_n close to the design value, which preserves handling balance and suppresses resonant peaks (Wang et al.; Soliman and Kaldas). Air suspension primarily varies mainly k, directly shifting ω_n—enabling adaptive retargeting of ride frequency to improve comfort under changing loads or to stiffen handling upon demand (Moheyeldein et al.). Hydraulic systems have the potential to vary both k and c and exert net forces, shifting eigenvalues and governing modal coupling and damping across the vehicle, achieving the broadest performance envelope at a cost and complexity penalty (Guo et al.; Ling et al.).

These distinctions highlight different control strategies which suspension designers can choose among depending upon their goals such as cost, weight, or reliability. When maintenance of a desired ride frequency is a top priority and suppressing resonant peaks without changing stiffness is required, MR is frequently employed. When ride frequency is required to adapt automatically to changing loads, air suspension provides a direct and efficient solution by shifting k. When maximum possible reduction of perceived motion and maximum flexible control of body modes are desired, hydraulics offer unmatched capability but are costly to engineer because of their associated compromises. Overall, MR systems primarily regulate damping, air suspension systems shift stiffness, and hydraulic systems govern both.

Works Cited

Bica, Ioan, et al. "Physical Characteristics of Magnetorheological Suspensions and Their Applications." Journal of Industrial and Engineering Chemistry, vol. 19, no. 2, Mar. 2013, pp. 394–406, https://doi.org/10.1016/j.jiec.2012.10.008. Accessed 19 Nov. 2025.

Blevins, Robert D. Formulas for Natural Frequency and Mode Shape. Van Nostrand Reinhold, 1979, pp. 1–25, archive.org/details/formulasfornatur0000blev/page/n5/mode/2up. Accessed 19 Dec. 2025.

Bucchi, Francesco, and Basilio Lenzo. "Analytical Derivation and Analysis of Vertical and Lateral Installation Ratios for Swing Axle, McPherson and Double Wishbone Suspension Architectures." Actuators, vol. 11, no. 8, Aug. 2022, p. 229, https://doi.org/10.3390/act11080229. Accessed 29 Sept. 2025.

Gillespie, Thomas D. Fundamentals of Vehicle Dynamics. Society of Automotive Engineers, Inc, 1992, pp. 90–140, lic.haui.edu.vn/media/CN%20Oto/Fundamentaln%20of%20Vehicle%20Dynamics.pdf. Accessed 29 Sept. 2025.

GlobalSpec. "Convoluted Air Spring; Rolling Lobe Air Spring." GlobalSpec, 2026, www.globalspec.com/learnmore/mechanical_components/springs/air_springs. Accessed 10 Feb. 2026.

Goodarzi, Avesta, and Amir Khajepour. Vehicle Suspension System Technology and Design Synthesis Lectures on Mechanical Engineering. Synthesis Lectures in Mechanical Engineering, 2017, dacemirror.sci-hub.se/journal-article/d120391fcf48a173b02a7a4ca6d88902/goodarzi2017.pdf. Accessed 29 Sept. 2025.

Guo, Sijing, et al. "Hydraulic Integrated Interconnected Regenerative Suspension: Modeling and Characteristics Analysis." Micromachines, vol. 12, no. 7, June 2021, p. 733, https://doi.org/10.3390/mi12070733. Accessed 30 Sept. 2025.

Juvinall, Robert C., and Kurt M. Marshek. Fundamentals of Machine Component Design. John Wiley & Sons, 2012, pp. 110–35, gearkade.com/Gearkade%20content/Books/Fundamentals%20of%20Machine%20Component%20Design.pdf. Accessed 20 Sept. 2025.

Karnopp, Dean. Vehicle Dynamics, Stability, and Control. CRC Press, 2016, pp. 45–75, api.pageplace.de/preview/DT0400.9781466560864_A37764586/preview-9781466560864_A37764586.pdf. Accessed 20 Sept. 2025.

Ling, Qihui, et al. "Performance Evaluation and Analysis for Hydraulic Active Suspension Using the Collaborative Control Method." Measurement and Control, vol. 58, no. 6, SAGE Publishing, Sept. 2024, https://doi.org/10.1177/00202940241281454. Accessed 30 Sept. 2025.

MacPherson, Earle S. "US2624592A - Vehicle Wheel Suspension System." Google Patents, Patent Illustration, 21 Mar. 1947,

patents.google.com/patent/US2624592A/en?q=(macpherson+strut)&oq=macpherson+strut&sort=old. Accessed 19 Nov. 2025.

Maria, Phillip Di. “US8573605B2 - Active Suspension System and Hydraulic Ram Therefore.” Google Patents, Patent Image, 16 Feb. 2010, patents.google.com/patent/US8573605B2/en?q=(hydraulic+active+suspension)&oq=hydraulic+active+suspension. Accessed 19 Nov. 2025.

Meirovitch, Leonard. Fundamentals of Vibrations. Mcgraw-Hill, 2001, pp. 85–120, www.iust.ac.ir/files/fnst/ssadeghzadeh_52bb7/files/EB__Fundamental_of_Vibration.pdf. Accessed 19 Dec. 2025.

Moheyeldein, M. M., et al. “An Analytical Study of the Performance Indices of Air Spring Suspensions over the Passive Suspension.” Beni-Suef University Journal of Basic and Applied Sciences, vol. 7, no. 4, Dec. 2018, pp. 525–34, https://doi.org/10.1016/j.bjbas.2018.06.004. Accessed 30 Sept. 2025.

Mordor Intelligence. “Automotive Suspension System Market | 2022 - 27 | Industry Share, Size, Growth - Mordor Intelligence.” Www.mordorintelligence.com, 25 Feb. 2026, www.mordorintelligence.com/industry-reports/automotive-suspension-system-market. Accessed 2 May 2026.

Nayfeh, Ali Hasan, and Dean T. Mook. Nonlinear Oscillations. Wiley, 2007, pp. 15–35, onlinelibrary.wiley.com/doi/book/10.1002/9783527617586#aboutBook-pane.

Rajamani, Rajesh. “Vehicle Dynamics and Control.” Mechanical Engineering Series, Springer US, 2012, pp. 160–220, https://doi.org/10.1007/978-1-4614-1433-9. Accessed 29 Sept. 2025.

Rao, Singiresu S., and Philip Griffin. Mechanical Vibrations. Harlow Pearson, 2018, pp. 140–85, api.pageplace.de/preview/DT0400.9781292178615_A31460862/preview-9781292178615_A31460862.pdf. Accessed 19 Dec. 2025.

Soliman, AMA, and MMS Kaldas. “Semi-Active Suspension Systems from Research to Mass-Market – a Review.” Journal of Low Frequency Noise, Vibration and Active Control, vol. 40, no. 2, Oct. 2019, pp. 1005–23, https://doi.org/10.1177/1461348419876392. Accessed 30 Sept. 2025.

Tapia, Pablo, et al. “Comparative Analysis of MacPherson and Double Wishbone Suspensions for an Electric Off-Road Vehicle Retrofit.” World Electric Vehicle Journal, vol. 16, no. 4, Multidisciplinary Digital Publishing Institute, Apr. 2025, pp. 228–28, https://doi.org/10.3390/wevj16040228. Accessed 29 Sept. 2025.

Thomson, William. Theory of Vibration with Applications. CRC Press, 2018, pp. 190–225, api.pageplace.de/preview/DT0400.9781351408509_A35171716/preview-9781351408509_A35171716.pdf. Accessed 9 Dec. 2025.

Wang, Zunming, et al. “Advancements in Semi-Active Automotive Suspension Systems with Magnetorheological Dampers: A Review.” Applied Sciences, vol. 14, no. 17, MDPI AG, Sept. 2024, p. 7866, https://doi.org/10.3390/app14177866. Accessed 30 Sept. 2025.

White, Wil. Preliminary Investigation Natural Frequencies of Pickups and Medium and Heavy Trucks for Use in Component Design. 19 July 2018, dot.ca.gov/-/media/dot-media/programs/research-innovation-system-information/documents/preliminary-investigations/truck-natural-frequency-pi-a11y.pdf. Accessed 29 Sept. 2025.

Wong, J. Y. Theory of Ground Vehicles. John Wiley & Sons, Inc, 2022, pp. 245–75, onlinelibrary.wiley.com/doi/book/10.1002/9781119719984. Accessed 19 Dec. 2025.

Yartsev, Alex. "Damping." *Deranged Physiology*, 20 Jan. 2018, derangedphysiology.com/main/cicm-primary-exam/cardiovascular-system/Chapter-734/resonance-damping-and-frequency-response. Accessed 19 Nov. 2025.

Psychodynamic Psychotherapy In an Empirical World: A Literature Review of Clinical and Neuroscientific Efficacy By Gursher Chopra

Abstract

This literature review examines the clinical and neuroscientific validity of psychodynamic psychotherapy in an increasingly empirical psychological landscape. While modern curricula often favour Cognitive Behavioural Therapy (CBT) for its short term metrics, meta analysis reveal that psychodynamic therapy interventions yield superior long term personality restructuring and lower relapse rates due to the "sleeper effect". Utilising the framework of dual-aspect monism, this paper bridges objective neuroimaging and neurochemistry with subjective lived experiences. Classical psychodynamic concepts such as John Bowlby's internal working models, Mary Ainsworth's attachment styles and Sigmund Freud's theories of transference and repetition compulsion are mapped directly onto modern trauma neuroscience, implicit processing and sensory gating mechanisms. Finally Carl Jung's collective unconscious and archetypes are integrated with modern evolutionary psychology with cognitive schemas. Ultimately, this review demonstrates that psychoanalysis provides a necessary, holistic framework for understanding human suffering that rigid laboratory metrics cannot replicate.

Introduction to psychodynamic

Therapy in the modern world

In the rapidly empirical world we live in, Psychodynamic therapy is viewed as an abstract anomaly in psychology as most psychology students are taught Freud in the most shallow possible way and are then forced to learn his most unlikely theories which either leads students to dislike the subject and likely not keep It as a major or at all in further studies. Students who retain an interest often develop an extremely empirical, literalistic and statistical view of psychology. This is a highly reductionistic view of an abstract science that, through rigorous research, has been shown to encompass both the physical and the metaphysical but because of the school systems, Freud becomes a hated part of the curriculum and other important Psychologists such as Carl Jung, Anna Freud, John bowlby etc are forgotten or ignored because their work dabbled in the "unprovable" side of psychology, empirical researchers have shown that patients who go through Psychodynamic therapy consultations are more likely to not only return to the normal mental state but are proposed to have a "sleeper effect" (Shedler, 2010) in which the therapy helps to improve quality of the mind and emotions unconsciously and helps patients become more emotionally mature for future life challenges. Psycho-dynamic therapy has been statistically proven to be a much more personal, inclusive and metaphysical form of a Psychological treatment process even the patients of CBT(Cognitive behavioural therapy) but meta analysis indicates that while Cognitive Behavioural Therapy (CBT) provides robust short term symptom reduction, Psychodynamic interventions yield higher rates of long-term

personality restructuring, preventing symptom relapse post termination (Shedler, 2010) and have a transformative effect on the patient and their personal and professional relationships. Psychodynamic therapy is not restricted by statistics or empirical evidence and has a subjective and objective view on each patient which leads to a more inclusive and comfortable environment for a patient. In psychodynamic therapy patient's past is studied and their mental "shackles" are pinpointed in trauma,PTSD and panic attack patients after which by the use of psychodynamic therapy, those shackles are destroyed and the patients are independent and live more fully in the present instead of living under the tyranny of the past.

The mind and the brain represent a dual-aspect monism: a single entity viewed from two distinct perspectives (inside-out vs outside-in) (Solms, 2015)

It is a justifiable statement that while there is no mind without the brain, a brain can exist without a mind. We cannot place the mind on a physical scale, yet we can study it comprehensively, it is a profound scientific conundrum. Neuroscience has been using many methods to study our brain such as our anatomy, neurochemistry and fMRIs. Anatomy by studying the macroscopic and microscopic structure of the brain. Neurochemistry by focusing on the molecular level of our neurotransmitters and the hormones facilitating through them when neurons communicate with each other. In this process they have found the theorised neural These activities which are quantifiable and of well practiced scientific method of study, but our mind isn't pathways for schizophrenia and autism. Meanwhile, fMRI measures brain activity by detecting changes in blood oxygen flow, offering researchers an outside-in perspective on the brain. In such a fashion. We can take predictions or better known as hypothesis and apply scientific research methods to it and then a theory is published after predicted results or hypothesis is proved right or wrong by applying the scientific method. Even then we have a small error percentage which is the dependent variable which cannot be controlled. Therefore are not factored because they cannot be. Which is why the application of the psychodynamic perspective and theory is filling the tiny slits even science cannot account for neither fill because psychodynamic therapy is subjective and objective not only objective in comparison to other psychological theories. This exact aspect gives it the edge for a more well rounded perspective that a psychologist must consider for more well rounded therapy sessions. Psychoanalysis looks at the brain from the inside-out (such a person's subjective lived experiences, emotions, outbursts, dreams). Neuropsychoanalysis (Solms, 2015) acts as a bridge connecting them as the other therapy techniques. Look at the brain from the outside-in connecting them, as other therapy techniques look at the brain from the outside-in and psychodynamic therapy looks at it from the inside-out but together as Neuropsychoanalysis. It looks at it from both perspectives and especially when the subject is as important as the human mind, the tool that has led to the discovery of the theory of relativity, gravity, calculus and nuclear weapons. An instrument of this complexity should not be studied restrictively behind rigid empirical barriers. All theories should be welcome which could make for a more stronger understanding of the mind and maybe further our thinking ability to be abstract and not rigid.

Psychoanalytic methods that oppose physicalism and reductionism

A psychological symptom occurs when an itemized prediction fails to meet a biological need. The unmet need triggers raw, conscious emotional pain (Solms, 2015) from the brain stem (the Id) because the prediction is deeply buried in unconscious memory, the patient experiences emotional distress without understanding why. The psychoanalytic methods such as free association(where a person speaks aloud whatever thoughts, words or images enter their mind, completely uncensored and without filtering for logic or relevance.) and exploring interpersonal environments(involves taking a step back to evaluate how you interact with others and how those people impact your emotional well-being, behaviours and energy. It is a concept often used in counselling and personal development). These techniques for uncovering the subjective experience of feelings which standard neuroscience often ignores. Psychoanalysis opposes outdated physicalism or reductionism by arguing that subjective experiences(the psyche) cannot fully be explained by functional brain scans alone (Solms, 2015). Psychoanalysis provides theories through sophisticated vocabulary and the necessary infrastructure to understand complex, subjective meanings behind biological effects. We will further examine these in the paper.

Attachment, defence mechanisms and child care in modern psychodynamic therapy (Bowlby, 1979)

Using internal working models(a cognitive framework essentially a mental blueprint that we use to understand the world, ourselves and our relationships). While these are highly resistant to change because they act as filters causing us to unconsciously seek out relationships that confirm our existing beliefs they are not set in stone (Bowlby, 1979). Deeply impactful experiences later in life, such as a long term secure romantic relationship or consistent therapy, can restructure an old, unhealthy model into what psychologists call an ‘earned secure’ attachment. These models explain how early relationships shape adult personality. Freud’s explanation of transference should be studied especially in today’s generation which is statistically the most depressed and suicidal generation out of all of them and the technique of transference (brain’s shortcut, It assumes that because someone reminds us of a figure from our past, they will behave exactly like them). This technique from psychoanalysis is very important to study in modern day psychology as some patients use their unmet parental needs either on to their therapist or teacher or any guiding figure. Some might transfer their issues on to another friend because of a physical trait they possess and such an abstract phenomena can only be studied in psychoanalysis not in rigid scientific methods. This is not to criticise scientific methods by no means instead it is to broaden their scale and allow some meta-physical aspects to enter the physical aspects and even though we cannot use the scientific method on transference but we can use our observational skills and admit, patterns of transference are quite visible in modern therapy practices.

Psychodynamic phenomena which can be observed in scientific laboratories (Ainsworth, 1969)

Psychoanalytic concepts and phenomena are observable in laboratory settings such as the famous 'strange situation' (Ainsworth, 1969) protocol in which early children's behaviour was studied by leaving them alone in a room and letting different people enter the room some that the child was familiar with such as his or her parents or a stranger with which the child was not familiar with and see the child's distress when their parent would leave and anxiety when they would be left alone with a stranger and finally their relief when their parent would come back. This study empirically showed the maternal-child attachment in a scientific environment and this behavioural study showed how important early caregiving to children is and how even the slightest mistake leaves a lasting effect on the child in the future on their psyche. This demonstrated that unconscious defences such as avoidant or anxious attachment styles can be scientifically observed and classified in laboratory settings. It also shows psychoanalytic defence mechanisms such as the unconscious exclusion of distressing thoughts, memories or impulses from the conscious mind (repression) and defence mechanisms that involve black and white or good and bad or dichotomous thinking (splitting). These are logical and biological responses to relational trauma and loss. Detachment is shown in a child who has lost their parents or children who grow up in foster homes or orphanages which is a visible representation of repression and the child shuts down their attachment system to protect themselves from overwhelming despair, providing a clear behavioural view of unconscious defence mechanisms and this shows not only scientific but practical application and relevance of psychoanalytic dogmas. Psychoanalysis takes the step outside the physical and enters the metaphysical which rigid science cannot.

The therapeutic efficacy of psychoanalysis (Shedler, 2010)

Early maternal care physically shapes infant brain development, alerting stress maintenance and cortisol receptor levels and how emotions are regulated within the mental processes whether in a healthy manner or an unhealthy manner and these aspects affect an individual's life as a whole and other databased psychological schools cannot explore such complexity in human behaviour and only psychoanalysis can and with psychoanalysis these untouched arenas of psychology are explored that the other empirical branches can't because of their rigidity.The efficacy of psychodynamic therapy lies in the fact that a psychodynamic therapist is not only just an analyst but a surrogate attachment figure which fixes previously done mental damage with the patient's environment or health or experiences etc. It provides a "secure base" (Bowlby, 1979) for patients and allows them to process the past and present in a safe and comfortable environment with the psychoanalyst. Allows patients to safely explore challenges and update their rigid internal working models with environments which foster growth not denouncement and nurture the patient with a stable relational environment.

Freud's scientific theories that combat modern cognitive psychology

Freud correctly predicted that the brain performs complex executive tasks without conscious awareness (Freud, 1915). Freud's research in discovering automatic, self-related mental processes or more such as biases, hidden habits or defence mechanisms (Freud, 1894) that operate entirely outside your immediate awareness. Concepts in modern psychology such as the cognitive unconscious (Kihlstrom, 1987) and implicit cognitive processing in such as these operations work they are merely advanced mechanisms which were earlier theorised by Freud but not taken seriously and since he didn't have the proper technological advancements, his theories could only remain theories but Freud predicted both of the modern psychological phenomena I have listed earlier before they were even discovered such as the unconscious ego, ID and defence mechanisms, projection etc (Freud, 1894-1923). He laid the groundwork on which the modern cognitive psychologists lay their theories and principles on and even phenomena such as repression or projection (Freud, 1894) are simply automated non-conscious mental operations used to avoid psychological distress. Even now with the medicinal advances such as anti-depressants. It is still proven that depression is at an all time high which means studying psychology fully and embracing all aspects and not being extremely rigid is the right attitude to such a wonderful and vast subject. This makes the argument that psychoanalysis is too subjective fall apart because cognitive and even genetic psychology is too subjective as well. Genetic psychology says that if a certain gene is in a person their behaviour can be predicted but what about identical twins which have 100% of the same genes yet have different experiences, mental processes and personalities. Even the twins differ on views which clearly shows that it's not the fact that psychoanalysis is too subjective or unprovable it is purely the fact that it stems from an epistemological bias that favours laboratory-controlled metrics over robust, ecologically valid qualitative outcomes and instead prefer the laboratory experiments which cognitive psychology uses but they forget that even psychoanalysis conducts experiments in laboratories.

The Modern view of psychodynamic therapy (Freud, 1914-1920)

The clinical reality of the phenomenon where a person unconsciously repeats traumatic, painful or distressing situations from their past. This repetitive compulsion (Freud, 1914) which was theorised by Freud aligns with modern PTSD and trauma neuroscience (Van der Kolk, 2014) The subcortical survival mechanisms were described by Freud who was denounced by modern cognitive psychologists and his theory re-named even though his ideas which had a connection to modern cognitive psychology like trauma bonding, re-enactment and the way unintegrated traumatic memories break down the brain's cognitive processing systems. Freud defined that trauma occurs when an environmental shock is so sudden or massive that it pierces the protective shield, flooding the mind with anxiety while modern neuropsychology states that under extreme stress or PTSD, sensory gating collapses. The brain loses its ability to suppress redundant stimuli, forcing the individual to experience memories and current surroundings with an agonising, unfiltered vividness. Such similarity in core definitions cannot be ignored and more over psychoanalysis doesn't just stick to one portion or characteristic of the brain instead it is a holistic discipline that refuses to oversimplify human sufferings for the sake of easy metrics.

Freud proposed that human psychology is a constant tug of war between two opposing primal forces: Eros (the life drive) and Thanatos (the death drive). Standard behavioural psychology often assumes humans naturally strive for rewards and avoid pain. Freud used the friction between Eros and Thanatos to explain why people consciously or unconsciously repeat harmful patterns, such as staying in toxic relationships, succumbing to addiction or self-sabotaging their own success. The note argues that by acknowledging this dark, irrational side of human nature, psychoanalysis provides a much more complete and realistic framework for deep human suffering than basic behavioural models that rely purely on measurable habits or external stimuli.

Conclusion: Jungian theories in the modern world

The collective unconscious (Jung, 1916) was the psychological expression of our shared human genome and evolutionary psychology and behavioural genetics. Even birds are born with the innate neural wiring to build nests without being taught, humans are born with inherited, subcortical brain structures that shape how we perceive reality, experience emotions and form social bonds. Carl Jung argued that “archetypes” (Jung, 1919) are universal patterns or symbols inherited from our ancestors. Jung’s archetypes connect to modern “schemas” , the brain’s built-in filing system used to quickly sort vast amounts of data without wasting energy. Jung argued that humans are genetically hardwired to recognise certain threats instantly (like spiders or angry expressions) for survival. Archetypes are empty, innate hardware templates. Your personal upbringing and culture then act as the software that fills them in. Jung came up with the idea of the “shadow” (Jung, 1912) which refers to the hidden, repressed or dark parts of our personality. When a person has a trait or urge that contradicts who they think they are, it creates intense mental pain known as cognitive dissonance (Festinger, 1957) to mitigate this pain, the brain hides these traits in the shadows to protect the ego. The mind cannot tolerate seeing these hidden traits in itself, it automatically projects them onto other people. For example, someone who suppresses their own anger might constantly accuse others of being hostile. This proves that the unconscious shadow is not passive. It actively warps daily social interactions, relationships and judgements.

Works Cited

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. American psychologist, 65(2), 98-109

Solms, M. (2015). The conscious Id. Neuropsychoanalysis, 15(1), 81-119

Ainsworth, M.D.S. and Wittig, B.A. (1969). Attachment and exploratory behaviour of one year olds in a strange situation. In B.M.Foss (Ed.), Determinants of infant behaviour (Vol. 4, pp. 111-136). London: Methuen.

Bowlby, J. (1979). The Making and Breaking of Affectional Bonds. London: Tavistock.

Freud, S. (1894). The Neuro-Psychoses of Defence. Standard Edition (Vol.3).

Freud, S. (1914). Remembering, Repeating and Working-Through (Further Recommendations on The Technique of Psychoanalysis II). Standard Edition (Vol.12).

Freud, S. (1915). The Unconscious. Standard Edition (Vol.14).

Freud, S. (1923). The Ego and the Id. Standard Edition (Vol.19).

Jung, C. G. (1912). Psychology of the Unconscious. New York: Moffat, Yard and Co.

Jung, C. G. (1916). The Concept of the Collective Unconscious. Collected Works (Vol.7).

Jung, C. G. (1919). Instinct and the Unconscious. Collected Works (Vol. 8).

Festinger, L. (1957). A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press.

Kihlstrom, J.F. (1987). The Cognitive Unconscious Science, 237(4816), 785-794.

Van der Kolk, B.A. (2014). The Body Keeps the Score: Brain, Mind and Body in the Healing of Trauma. New York, NY: Viking.

Freud.A. (1936). The Ego and the Mechanisms of Defence. London: Hogarth Press

Nero and the Great Fire of Rome, 64 CE By Lily Li

Introduction

On a hot July night in 64 CE, a small fire erupted among the wooden stalls of the Circus Maximus, Rome's largest chariot-racing stadium. Over the next nine days and nights, it rapidly expanded and consumed temples, homes, and entire districts, as "wailings of [the] terror-stricken" filled the streets, leaving ten of Rome's fourteen regiones (districts) in ruins (Tacitus 15.38). The scale of destruction was unlike anything the city had faced before. In the aftermath, rumors spread that the emperor at the time, Nero, had ordered the fire and sang from his palace as the city burned, though some also reported that he directed relief efforts and provided shelter. What started as an ordinary city fire quickly made Nero one of the most infamous rulers in Roman history.

As destructive as the Great Fire was to Rome, so too did it prove ruinous to Nero's legacy. Recent studies by archaeologists, environmentalists, and historians agree that Nero likely played a limited role in the outbreak of the fire (Romm). Rome's architecture, consisting of densely packed wooden tenements vulnerable to hot southeasterly winds, was the true cause of the fire's widespread destruction. Why, then, did Nero gain his reputation as Rome's arsonist? How did ancient Roman historians—namely Tacitus, Suetonius, and Cassius Dio—portray Nero's motives and behaviors surrounding the Great Fire, such that his blame endured for nearly two millennia?

To answer these questions, it is necessary to examine the sources that established Nero's legacy. There are only three surviving accounts of the Great Fire of Rome, written by Tacitus, Suetonius, and Cassius Dio. These were composed between approximately 117 and 229 CE—at least fifty years after the Great Fire—but they remain key evidence for the event. Other contemporary sources, including Vitruvius, Pliny the Elder, Seneca, and Josephus, offer no direct commentary on the fire but provide valuable insight into Rome's architecture, social structure, and imperial politics during the Julio-Claudian period. In addition, archaeological evidence has clarified aspects of the city's layout and the extent of the fire's destruction. While modern scholars like Miriam Griffin and Edward Champlin offer influential biographies of Nero's life, they analyze the fire within larger social and cultural contexts without specifically focusing on Nero's involvement. This paper builds on their work by bringing together ancient literary accounts and selecting archaeological evidence to better understand how Nero's role in the fire was interpreted, recorded, and remembered.

This paper argues that, while Nero likely played a limited role in the fire's outbreak, his image as Rome's arsonist endures because of his controversial personality and the self-interests of ancient historians. By analyzing the motives behind these portrayals, this study demonstrates that the persistence of the arsonist story speaks more to the priorities of ancient historians than to Nero himself. Tacitus, the most acclaimed Roman senator of his time, used the fire to critique imperial power and idealize Rome's early Republic. Suetonius, an imperial secretary, used gossip to portray Nero as a corrupt emperor. Cassius Dio, writing more than a century later, depicted the

fire as a moral lesson about the consequences of autocratic rule. The story of Nero fiddling as Rome burned—the image most associated with his reign—likely came from ancient historians' own imaginations and interpretations rather than Nero himself.

The Great Fire of 64 CE

Among the dangers that threatened daily life in imperial Rome, fire was the most constant and destructive. The acclaimed Roman poet Horace once mocked the anxieties of the rich in his Satires, "Terrified night and day of thieves or fire or slaves who rob" (Horace 1.1.77). The fear of fire was not unfounded. The city's architecture had long been prone to catastrophe. Vitruvius, an architect under Augustus's reign, warned that Rome's insulae, tenement buildings made of timber and rising several stories, made the city liable to frequent and very great fires (Vitruvius 2.8.20). The poet Juvenal, with characteristic fatalism, later described the capital as "held up for the most part by slender props," a city built for collapse, whether due to fire or time (Juvenal 3.193-194). Although ancient historians would eventually credit Nero's sadism for the fire, evidence from both archaeologists and Roman contemporaries highlights Rome's early vulnerability to a devastating fire. In July 64 CE, that threat became a reality.

The fire began, according to the Roman historian Tacitus, "in that part of the circus [Maximus] which adjoins the Palatine and Caelian hills" (Tacitus 15.38; see Fig. 1). Tacitus, whose Annals remains the most reliable account of this event due to its relative objectivity, described the fire as spreading with astonishing rapidity. A dry southeast wind called the Sirocco fueled the fire, which caused a minor emergency to worsen into a conflagration. The Circus Maximus—an area dense with shops and workshops packed with flammable goods—ensured that, once the flames began, containment became virtually impossible. The fire raged unchecked for six days and seven nights. It was briefly quelled only to flare up again, continuing for three more days (15.40).

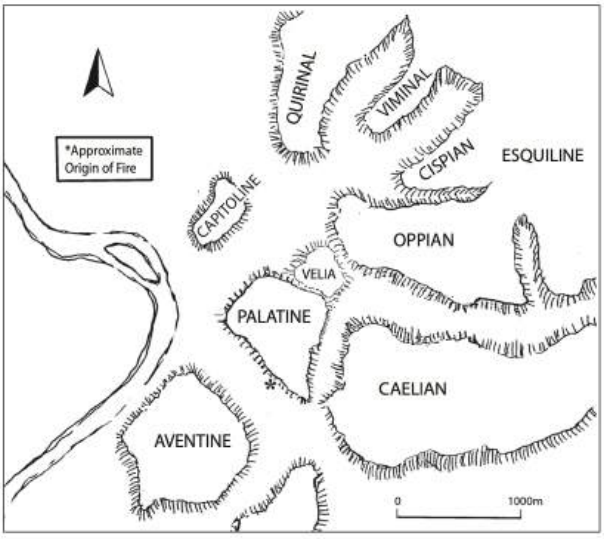


Fig. 1: The hills of Rome. The asterisk (*) indicates the origin of the fire at the base of the Palatine and Caelian hills, as described by Tacitus. Source: Louis, A.

The scale of destruction was nearly unprecedented. Three of Rome's fourteen regions were "levelled to the ground," and seven more were left with "only a few shattered, half-burnt relics of houses" (Tacitus 15.40; see Fig. 2). Only four survived largely intact. The fire swept eastward from the Circus Maximus over the Palatine and Caelian, south to the Aventine, and north to the Esquiline, consuming neighborhoods with terrifying speed and unpredictability. Suetonius, writing from the perspective of an imperial secretary, described how many perished, overtaken by the flames and noted that survivors took refuge in "tombs and monuments," imagining themselves already among the dead (Suetonius Nero 38). These accounts reflected the widespread confusion and fear during the fire.

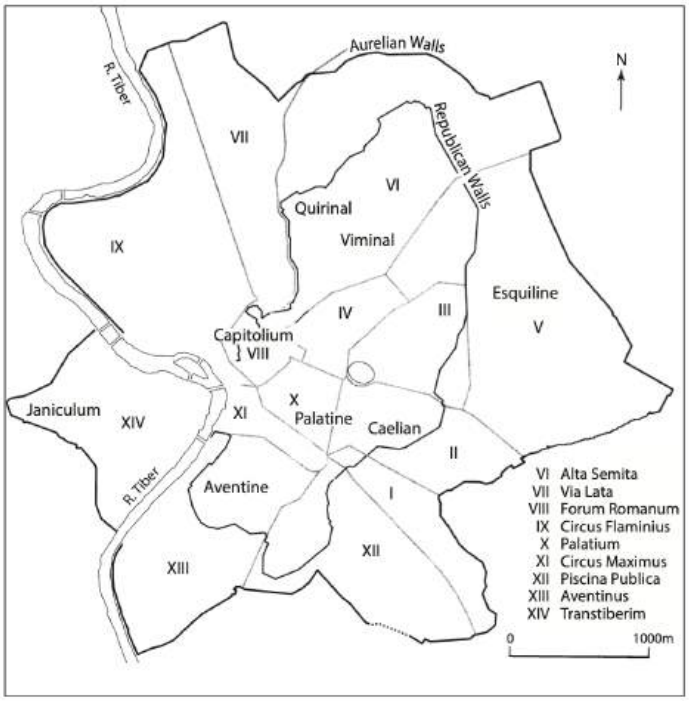


Fig. 2: Map of Rome's fourteen regions. Source: Skinner, J.

The fire's consequences were felt in every level of Roman society. Archaeological evidence, consisting of burnt layers on the Palatine and Aventine as well as damaged masonry near the Forum, confirms not only the scale but also the geographic reach of the destruction (Gualandi 75). Among the buildings lost were the Temple of Luna, the House of the Vestals, the shrine of Jupiter Stator, and sections of the imperial palace. Public archives, commercial warehouses, and countless private homes were reduced to ruins. This meant that many of Rome's official records were lost, resulting in gaps in the city's historical and administrative memory.

The number of casualties remains unknown, partly because so little documentation survived the fire. The chaos of the event, the lack of systematic record keeping, and the subsequent destruction of archives ensured that the loss of life, though certainly considerable,

was never formally tallied. What is certain is the scale of displacement: tens of thousands were rendered homeless, forced into makeshift camps in public spaces or the emperor's own gardens.

The fire's origins were the subject of immediate speculation. As the city burned, rumors circulated that the emperor himself had played a role. It was said that while Rome burned, Nero ascended to his private stage and “sang of the destruction of Troy” (Tacitus 15.39). Some accused imperial agents of orchestrating the fire, or at least impeding efforts to extinguish it. Tacitus noted in his Annals that "all human efforts, all the lavish gifts of the emperor, and the propitiations of the gods, did not banish the sinister belief that the conflagration was the result of an order" (15.44). The suspicion that the disaster was not entirely accidental would shape Nero's legacy for centuries.

Despite the above, modern historians now agree that Nero’s guilt is unlikely. According to the accounts of ancient historians, Nero was staying at his villa in Antium, some thirty-five miles from Rome, at the moment the flames erupted (Barrett 72). Tacitus, himself a severe critic of Nero, recorded that the emperor did not return to Rome until news of the fire reached him. When he did return, Nero took action to alleviate suffering. He threw open “the Campus Martius… and even his own gardens” to shelter the homeless and “raised temporary structures” for those left without shelter (Tacitus 15.39). Food and grain were imported at the state’s expense, and relief measures were organized at imperial command. To better protect the city from future fires, Nero’s reconstruction program introduced new regulations requiring greater street width and the use of fire-resistant building materials like brick and stone.

Yet, public suspicion remained for several reasons. The emperor’s decision to appropriate large areas of cleared land for the construction of his lavish palace, the Domus Aurea, was seen as a display of imperial self-indulgence (Griffin 137). Additionally, Nero attempted to divert blame for the fire by accusing the minority Christian community of arson. The resulting persecution of Christians caused public unease, given the severity and brutality of the punishments.

The Great Fire of 64 CE caused widespread destruction and left a lasting impact on Rome. While ancient accounts raised suspicions about Nero’s involvement, modern historians generally agree that the fire was an accident, exacerbated by the city’s structural vulnerabilities and environmental conditions. Given Nero’s absence from Rome at the time, his documented relief efforts, and the lack of credible evidence linking him to the disaster, it is highly likely that his role in starting the fire was minimal.

The Controversy Surrounding Nero

In the aftermath of the Great Fire, the cause of the disaster led to immediate speculation. The crisis demanded an explanation, and in the anxious days that followed, suspicion began to circle the emperor. The familiar story of Nero playing the fiddle while the city burned was a later invention; the instrument did not yet exist in 64 CE, and even the earliest sources mention only that Nero was rumored to have sung verses about Troy, a detail that Tacitus treated with skepticism (Tacitus 15.39). Still, after nearly two thousand years, the rumors have not faded.

Why did Nero's name become attached to the burning of Rome? The answer lies in part in the enduring legacy of Nero's controversial personality. Nero's reputation for theatricality and cruelty made him a convincing villain in the eyes of ancient writers and later generations. His character helped make the story endure.

The world that Nero inherited was shaped by the stern ideals of the Roman elite. Since the reign of Augustus (27 BC-14 CE), emperors had been expected to embody the virtues of restraint, dignity, and military command, presenting themselves as guardians of order and tradition (Brunt and Moore 5). The Senate, composed of Rome's most distinguished families, looked for rulers who would govern with the gravitas of statesmen and the discipline of generals. Nero, however, challenged these expectations from the outset. He was drawn to the arts, performing publicly as a singer, poet, and even charioteer, roles that were seen as inappropriate for a Roman aristocrat, let alone an emperor. Suetonius reported that Nero competed in public as a lyre-player, and even "sang tragedies" to the dismay of the senatorial class, who regarded the behavior as humiliating and dangerous to the dignity of the imperial office (Suetonius Nero 21). Dio wrote of the Senate's growing discomfort, suggesting that Nero's behavior was far beyond any remonstrance (Dio 62.11). For many elites, his open embrace of theater and performance moved too far from the traditions established under Augustus and Tiberius. Yet the response among the common people of Rome was, in the early years of Nero's reign, different. When he was first proclaimed emperor, Nero's generosity and lavish public entertainment won him widespread popularity. Historian Anthony Barrett characterizes him as "Rome's Golden Boy," whose early favor among the urban masses suggested potential for a promising reign (Barrett 10).

Nevertheless, Nero's early popularity quickly declined after he targeted members of his own family. Britannicus, the only son of Emperor Claudius and his legitimate heir, was, by birth, a standing challenge to Nero's position. Recounting the events of 55 CE, Tacitus reported how Britannicus was served a poisoned drink at a banquet, which "so penetrated his entire frame that he lost alike voice and breath" (Tacitus 13.15). This murder was also documented by other historians, including the Roman-Jewish scholar Josephus (Josephus 20.8). In all accounts, the abrupt and public nature of the boy's death fed immediate suspicion that Nero had orchestrated the murder to secure his hold on power. In fact, Tacitus explicitly stated that Nero recruited a known assassin, Locusta, to poison Britannicus, afterwards rewarding her with pardons and official favor (Tacitus 13.15). Later, in 59 CE, Nero arranged the assassination of his mother, Agrippina the Younger, whose political influence threatened his authority. Dio described how Nero first attempted to kill her by sabotaging a boat designed to collapse at sea; when that failed, Nero ordered her execution at her villa under the pretense of suicide (Dio 62.12). The horror of a son murdering his mother, particularly as she had helped secure his accession, demonstrated how Nero's ambition disregarded many of the sacred Roman values. Together, the two assassinations made Nero a dangerous figure in his pursuit of power.

In a further display of personal cruelty, Nero targeted his politically benign wife, Octavia. Unlike Britannicus, Octavia posed no threat to Nero. She was young and popular with the

Roman people, who saw her as a symbol of continuity and innocence. Nevertheless, in 62 CE, after Nero accused Octavia of adultery on questionable grounds, he executed her by opening her veins and presented it as suicide. The emotional weight of her death was captured in the pseudo-Senecan tragedy Octavia, reflecting the public mood: "Oh! ye kind breezes, convey away this victim, Octavia, far away, from any cruel punishment, I pray, to the temples of Trivia" (Pseudo-Seneca 982). The play portrayed Octavia as an innocent and tragic figure and described prayers offered for her to find mercy in the afterlife. It represents a rare indication of sympathy and compassion in ancient literature. Although Nero enjoyed popular support in the early years of his reign, the murder of Octavia was one of several turning points that gradually provoked resentment among the Roman public. By eliminating his harmless and popular wife, Nero developed a reputation for brutality in the period leading up to the Great Fire.

By the middle years of Nero's reign, imperial Rome was rife with fear and paranoia. Executions and forced suicides were routine forms of governance. The emperor's suspicion fell on political rivals, as well as senators, nobles, and former allies, often based on the vaguest charges of conspiracy or disloyalty. Tacitus recounted how accusations, which were sometimes fabricated by delatores (informers) hoping for imperial favor, could lead to swift deaths or demands for suicide. This was the case with the respected senator Thrasea Paetus and the Stoic philosopher Seneca (Tacitus 15.63). Similarly, under Nero's reign, the imperial court grew increasingly oppressive. He was known to blacklist victims, sometimes without a trial, and his informants reported private conversations spoken in the supposed safety of people's homes. Dio observed that under Nero, no one dared to speak or listen freely, and public life was an exercise in self-censorship and fear (Dio 62.15). Nero also treated punishments and executions as forms of entertainment. In fact, Suetonius recorded that Nero would order the most elaborate and sadistic punishments for minor slights or even just for his own amusement (Suetonius Nero 35). Meanwhile, the emperor's staging of ever more lavish games, his public performances as charioteer and singer, and his costly tours to Greece strained state finances to the breaking point. To fund his pleasures, Nero resorted to new taxes and forced confiscations. Thus, along with Octavia's murder, Nero's extravagant and oppressive policies shifted the public's view of him from admiration to growing contempt. Roman society increasingly came to see him as a cruel and tyrannical leader.

Nero was also publicly known for upending the most fundamental expectations of Roman society. In perhaps the most striking example, he entered a formal marriage with the freedman Pythagoras in early 64 CE, assuming the role of bride. Tacitus recorded that Nero "marr[ied] himself to one of that filthy herd, by name Pythagoras" in the presence of witnesses and that all the customary ceremonies were observed as if Nero were a woman (Tacitus 15.37). In Roman society, ideals of masculinity centered on self-mastery, public dignity, and the maintenance of clear distinctions between male and female roles. To occupy the passive or feminine position was widely regarded as incompatible with the status of a freeborn man, let alone an emperor. Dio likewise noted the open display of the marriage, reporting that Nero did not hesitate to appear in the role of wife before the Roman elite and to have the customary wedding ceremonies

performed (Dio 62.28). This act unsettled the ideas of gender, power and identity at the heart of Roman life.

So, in the aftermath of the Great Fire, Nero's past actions led to growing suspicion of his role in the disaster. While recovering from the fire, the public was eager to identify someone responsible. It was hard to believe that the disaster could be accidental. In practical terms, Nero's initial actions were not negligent. He was at Antium when the fire broke out but returned to Rome as soon as news reached him, personally directing relief efforts. Yet, as historian Stephen Dando-Collins notes, although these measures brought Nero "short-term popularity," the rumor of his guilt spread "even faster than the all-devouring flames" (Dando-Collins 12). In this context, Nero's cruel and theatrical persona made him an easy target for blame. Whether or not he was truly responsible, his reputation as a tyrant and an eccentric made his guilt plausible for the Roman public.

The construction of the Domus Aurea, Nero's luxurious palace, in the wake of the Great Fire, created the belief that Nero had exploited the disaster for personal gain. Before the fire, Nero had begun work on a more modest palace, the Domus Transitoria, but after the devastation cleared central Rome, he expanded his plans into the far more ambitious Domus Aurea. The Domus Aurea sprawled across the Palatine and Esquiline hills, featuring lakes the size of a small sea, gilded ceilings, and a rotating banquet hall powered by hidden mechanisms. Suetonius recorded that Nero's new residence was "ruinously wasteful," its grounds extending so far that the halls were "marked by milestones" and entire landscapes fashioned for imperial leisure (Suetonius Nero 31). The most infamous feature was the Colossus Neronis, a bronze statue of Nero that stood roughly one hundred feet tall, which was built as a monument to his own self-importance (Pliny the Elder 34). Although there is no definitive evidence that Nero orchestrated the fire to clear land for the project, the timing and scale of the Domus Aurea were politically unfavorable. Tacitus noted that people were homeless, hungry, and embittered while Nero began constructing a palace that "fool[ed] away an emperor's resources" (Tacitus 15.42). To both the Roman public and elite, the palace showed that Nero prioritized self-indulgence over Rome's recovery. The optics of the project, regardless of actual intent, confirmed views that Nero had ceased to govern as a dutiful princeps (leader).

Nero then attempted to deflect suspicion for the fire onto the generally unpopular Christian community. He ordered mass arrests and subjected those accused to elaborate and violent executions. Some were covered in animal skins and torn apart by dogs, others were crucified, and many were burnt alive to serve as living torches during night-time spectacles in Nero's gardens. These punishments were not just retributive but designed as public entertainment, with crowds gathering to witness the cruelty. The calculated theatricality of these events was noted by Roman observers. Juvenal, writing decades later, alluded to the spectacle of executions staged for popular amusement, while Tacitus suggested "a feeling of compassion" among the public for the Christians (Juvenal 1.155; Tacitus 15.44). Even among those indifferent or hostile to Christianity, the excess violence caused unease. The Roman elite and public began

to suspect that by blaming others and staging spectacles, Nero may have been attempting to divert attention from his own guilt or perceived shortcomings as Rome's leader.

The blame for the Great Fire thus became inseparable from Nero's name, as the disaster seemed to reflect the image he had created through his own previous actions. His violent spectacles, theatrical cruelty and disregard for tradition created a controversial personality that made him an enemy of the Roman order. The philosopher Seneca, once Nero's tutor, observed that a ruler's character influenced the public's judgment far more than any proclamation or defense (Seneca 1.1.4). In Nero's case, his image as a cruel and self-indulgent ruler led the public to judge him as Rome's arsonist, regardless of the facts. His infamous reputation made the fire's destruction seem consistent with his rule.

Tacitus's Annals

Of the three ancient historians who recorded the Great Fire, Tacitus is most renowned for the depth and influence of his account. Writing over fifty years after the event, he produced the most detailed surviving narrative, and his portrayal of Nero's role has influenced public memory ever since. Compared to the works of Suetonius or Cassius Dio, Tacitus's Annals (composed ca. 117 CE) is treated as the most reliable source. But his account of the fire demonstrates his personal perspective on power, morality and Rome itself. Tacitus's account of the fire was influenced by the values of his senatorial class: suspicion of autocracy, nostalgia for the early Republic, and belief in the moral decline of Rome. Though he never directly accused Nero of arson, Tacitus suggested doubt in a way that reflected his own anxieties about imperial rule. His account of the fire was thus carefully constructed to imply Nero's guilt.

Tacitus's historical perspective was heavily influenced by the trauma of the largely autocratic Flavian period. Although he completed the Annals during the reigns of Trajan and possibly Hadrian, Tacitus's political imagination was haunted by the rule of Domitian. Domitian's reign (81-96 CE) was characterized by fear, censorship, and the loss of senatorial authority. Trials for treason were common, and Domitian's government was known to execute senators under groundless charges. Tacitus referred to this era as one of "servitude," noting the moral and political degradation of the ruling class (Tacitus Agricola 2). After Domitian's assassination, Tacitus celebrated the return of civic life under Nerva, writing that the new reign marked "the rare happiness of times," when senators could think and express what they pleased (Tacitus Histories 1.1). He believed freedom of speech had been absent for years, and Nerva's succession improved the relationship between the Emperor and Senate. Modern historians have emphasized how deeply Domitian's rule influenced Tacitus's opinions. Classics scholar Miriam Griffin, for instance, suggests that Tacitus's works reflect a suspicion of power earned through experience (Griffin 234). Tacitus distrusted imperial authority because of the memory of a weakened Senate under Flavian rule, which in turn influenced his portrait of Nero.

Tacitus used his writing to express a longing for the lost values of the early Republic. Throughout his works, he presented liberty (libertas), dignity (gravitas), and moral authority (auctoritas) as the fundamental virtues of early Roman politics, values he believed had been

weakened by the rise of emperors. In the preface to the Annals, he lamented the end of open debate and the Senate's loss of influence, writing that after the early Republic, "there was not a vestige left of the old sound morality" (Tacitus 1.4). Tacitus believed that power corrupted both rulers and citizens, which diminished Roman values and freedom of speech. From his view, libertas, gravitas, and auctoritas referred to a political system where senators could speak openly, magistrates were accountable, and no single man could rule the state. In contrast to later periods of Rome, the early Republic had demanded discipline, duty, and service to the state. These ideals came across clearly in Tacitus's praise of politicians like Germanicus and Agricola, whom he presented as exceptions to the moral decay of the empire (Agricola 42). In recording history, Tacitus hoped to preserve republican virtues so that future generations might recognize what had been lost. His portrayals of Nero and the Great Fire should be read within the context of this larger goal: to document the collapse of a political tradition.

Tacitus's concerns about the moral decline of Rome influenced his judgments of emperors and their governance. He believed the imperial system had weakened the city's institutions and ethical foundations. In his Annals, Tacitus portrayed Rome's emperors as encouraging ambition, flattery, and obedience, while discouraging the older values of integrity and restraint. These were the values Tacitus associated with the senatorial class at its best: men who had once guided the Republic through civic duty and responsibility. Over the years, the principate (Rome's imperial government) favored silence and obedience, especially among senators who felt constrained by the limited freedom of speech. Tacitus often praised senators like Thrasea Paetus and Helvidius Priscus, who represented moral resistance to imperial rule. Thrasea notably walked out of a meeting in 59 AD when senators were pressured to congratulate Nero on the murder of his mother, Agrippina, while Helvidius openly criticized the emperor's policies (14.12). For Tacitus, these senators stood for a tradition of Roman senatorial leadership that had faded under emperors like Nero. Of Thrasea, Tacitus wrote that he remained firm in his "defiant independence" and loyal to the truth, even when the Senate became a place of submission (16.21). The Great Fire thus allowed Tacitus to demonstrate how, in his view, Nero had diminished senatorial authority. In the aftermath of the fire, the Senate played a limited role in the investigation or rebuilding of the city, while Nero attempted to divert blame and reshape the city on his own terms. For Tacitus, this suggested that Rome's republican foundations had been weakened.

In his account of the Great Fire, Tacitus presented Nero as a flawed emperor who encouraged imperial corruption. He acknowledged that the fire might have been accidental, but then immediately redirected attention to Nero's response. Though modern historians largely agree that Nero returned to Rome to direct relief efforts almost as soon as news of the fire reached him, Tacitus framed Nero's return as occurring only when "the fire approached his house" (15.39). This detail subtly brought into question Nero's priorities, suggesting that his first concern was his own property. Tacitus then described Nero's public response, including his financial aid initiatives and construction of shelters, but undermined this behavior by emphasizing that it "produced no effect" (15.39). With this phrase, he implied that Nero

attempted to act as a benefactor but failed to restore public trust. Tacitus noted that, despite Nero's actions, nothing "banish the sinister belief that the conflagration was the result of an order" (15.41). This phrase is revealing of Tacitus's writing strategy: he never directly accused Nero but aroused suspicion through the passive voice, using the phrase "sinister belief" without a named source and reporting that the fire "was the result of an order" without providing attribution. He thus implied uncertainty while encouraging the reader to draw a conclusion regarding Nero's guilt. This ambiguity allowed Tacitus to maintain the appearance of neutrality while guiding the reader toward moral judgment. He then noted that the fire began in the part of the Circus where men "openly hurled brands, and kept shouting that there was one who gave them authority" (15.38). Although the source of the authority is not explicitly named, the phrase implies that the men acted under Nero's commands**,** thereby distancing the claim from Tacitus while also engendering suspicion. Later, when describing Nero's reaction to the fire, Tacitus added that "a rumor had gone forth everywhere that… the emperor appeared on a private stage and sang of the destruction of Troy" (15.39). Tacitus did not offer evidence for this claim, but used the anecdote to imply a theatrical and irresponsible ruler. Toward the end of the account, he wrote that Nero's punishments of the Christians aroused "a feeling of compassion" because the crowd believed they were being destroyed "not, as it seemed, for the public good, but to glut one man's cruelty" (15.44). Here Tacitus relied on emotional detail to guide the reader's judgment without offering direct accusation. Yet his decision to include emotionally charged details suggests an interest in recording public feeling over the objective, historical truth.

Tacitus was the only ancient historian to describe the persecution of Christians in the aftermath of the Great Fire, which he may have used to criticize both Nero and the Christians. This event appears nowhere else in the historical record. Neither Suetonius, Dio, Pliny the Elder nor any other Roman historian mentioned it, despite their interest in Nero and the fire. Similarly, Christian writers Tertullian, Lactantius, Eusebius, and Jerome all identified Nero as the first persecutor of Christians, yet none of them connected the persecutions to the fire (Eusebius 2.25). Historian Barrett calls this silence "remarkable" (Barrett 164). The only ancient writer to mention the persecution was Sulpicius Severus, who wrote in the fifth century and openly relied on the Annals for his version (Severus 2.28). This makes Tacitus the sole source for what became the earliest instance of Christian martyrdom under imperial rule. To explain the silence elsewhere, archaeologist Charles Saumagne argues that the passage on this event in Tacitus's account was added later, identifying the references to the Christians and the punishments as interpolations (Saumagne 337-60). Others, like theologian Pierre Batiffol, contend that Tacitus confused two separate events (Batiffol 514). Both explanations share the same goal: to remove the persecution from the context of the fire, since no other evidence supports the connection. If Tacitus invented or distorted the event, he likely intended to denigrate both Nero and the Christians, an unpopular figure and an unpopular group at the time. Tacitus's claim that Nero "fastened the guilt" on the Christians conveyed to the reader that the charges were fabricated, though he never explicitly stated the Christians were innocent (Tacitus 15.44). In the same passage, he referred to the Christians as a "class hated for their abominations" and charged them

with “hatred against mankind” (15.44). These were common Roman attitudes toward Christians; yet, instead of challenging them, he allowed both the fabrication of the charges and the reputation of the victims to stand side by side. As Israeli historian Zvi Yavetz argues, Tacitus “did not want to clarify but to confuse the reader… to slightly incriminate both Nero and Christians, both of whom he hated” (Yavetz 183). Exaggerating the persecution and placing it in the aftermath of the fire allowed Tacitus to discredit both groups simultaneously, without fully endorsing either side. Through his account, Tacitus subtly directed the readers’ judgments of both the emperor and his scapegoats in a way that would influence public memory.

Tacitus’s account of the Great Fire reflected the concerns of his own time as well as the events of 64 CE. His story was carefully constructed, including deliberate gaps and suggestions, which invited the reader’s suspicion but not certainty. Consequently, Nero became a representation of everything Tacitus feared about imperial leadership. Harriet Flower, a professor of Classics at Princeton University, observes that Roman memory, which was written by the political elite, may have “distorted and embellished the past as often as it preserved it” (Flower 4). Tacitus's history fits perfectly within that tradition.

Suetonius’s The Twelve Caesars

Detailing the reigns of Julius Caesar and the first eleven emperors, Suetonius’s The Twelve Caesars (written ca. 121 CE) offered readers a new way to understand Roman rulers: through their private behavior, daily habits, and personal vices. His portrait of Nero was characterized by controversy and drama, and it remains one of the most enduring images of the emperor. Unlike the senator Tacitus, Suetonius worked as a secretary in the imperial court. His access to imperial records and court gossip allowed him to compile sensational and often salacious accounts primarily aimed at entertaining his readers. Suetonius’s portrayal of Nero was influenced by his background as a court insider, his interest in moral character over political structure, and his reliance on gossip to entertain elite audiences. The Great Fire of 64 CE therefore gave Suetonius material to craft a sensational story rather than to diligently document historical truth.

Suetonius’s position as a court official impacted both the content and tone of his writing. Unlike senatorial historians like Tacitus, Suetonius came from the elite equestrian (equites) class and served as magister epistularum (secretary of correspondence) under Emperor Hadrian. As writer Frank Cretella notes, this administrative role placed Suetonius in close proximity to the imperial court and gave him “access to documents written by earlier emperors” (Cretella 4). This position thus provided Suetonius with an insider’s perspective on the emperors he portrayed. Because his career depended primarily on imperial patronage and elite interest, he had greater freedom to concentrate on character, private life, and scandal, areas often ignored by traditional political or senatorial historians. However, stories focusing on these areas were based on rumor and court memory, giving his biographies intimacy and color that appealed to readers within the imperial world. His writing method therefore suited the tastes of his audience, who were educated but curious about the private lives of emperors.

In The Twelve Caesars, Suetonius organized his imperial biographies by theme rather than date. After a brief account of an emperor's lineage and path to power, he would shift to long sections that focused on physical appearance, speech, dress, sexual conduct, eating habits, and religious beliefs. Volume 4: Life of Caligula, for example, followed this model in full. One section detailed Emperor Caligula's cruelty, including an incident where he ordered an entire audience killed during a public show. Another focused on superstition. Suetonius described how Caligula claimed to speak with the sky god, Jupiter, and built a temple where he dressed as a god and accepted worship (Suetonius Nero 22). Later, he included reports of Caligula wearing women's clothing and parading through the palace (52). These examples were grouped thematically, not chronologically, to reinforce a particular image of the emperor. By collecting anecdotes about superstition and vanity, Suetonius wrote biographies designed to amuse and shock his audience. This method gave his work a lasting appeal.

Similarly, Volume 6: Life of Nero followed this structure closely. After a brief account of Nero's rise to power, Suetonius devoted much of the text to the emperor's private life. These sections focused on sex, violence, and theatricality. One of the most disturbing claims was that Nero "tried to turn the boy Sporus into a girl by castration, went through a wedding ceremony with him -- dowry, bridal veil and all… and treated him as a wife" (28). Suetonius added that Nero later married another man, this time playing the bride himself. The reliability of these stories is difficult to verify, as they relied on sources hostile to Nero and were written decades after his death. Nevertheless, Suetonius effectively portrayed Nero as a man who violated Roman expectations of gender and decency. For example, Suetonius reported that Nero "had relations with his mother" and later looked at her corpse with lust (34). He accused him of disguising himself and assaulting both men and women in the streets. Instead of providing a timeline or connecting these actions to Nero's policies as princeps, Suetonius presented them as individual and sensational anecdotes likely to emphasize Nero's personal behavior.

Furthermore, Suetonius introduced the Great Fire of Rome with limited effort to evaluate its cause or consider competing accounts. He wrote that Nero "watched the conflagration from the Tower of Maecenas" and sang about the fall of Troy in "tragedian's costume" (38). Later in the same section, Suetonius mentioned that Nero opened public buildings to shelter refugees and arranged food supplies. Yet he placed these actions next to reports that Nero sought the glory of founding a new city and naming it "Neropolis," creating doubt about Nero's intentions (55). Written nearly six decades after the fire, Suetonius's account was based on secondhand accounts characterized by suspicion and sensationalism; it thus lacked context and may be considered a sequence of anecdotes. Suetonius presented the fire as another example of Nero's indulgence and obsession with performance, consistent with the character built across the biography.

Suetonius and Tacitus took distinct approaches to the Great Fire. Tacitus introduced his account by stating the public belief that the fire "was the result of [Nero's] order," thus creating distance between the historian and the rumor (Tacitus 15.38). He proceeded to trace how the fire spread, how the city responded, and how Nero attempted to manage public opinion. Tacitus presented both public actions and rumors but separated them clearly. In contrast, Suetonius often

combined Nero's policies and harsh rumors, presenting them largely without commentary to enhance their dramatic effect. Suetonius also made limited effort to assess the credibility of his sources. For instance, in The Twelve Caesars, it is unclear whether the story of Nero singing was widely believed or merely circulated in hostile circles. The structure of Life of Nero supported this approach, whereby Suetonius documented events without extensively exploring their underlying contexts, consequently interpreting the fire as a reflection of Nero's broader eccentricities.

Suetonius then used the Great Fire to expose Nero's obsession with performance. Earlier in his reign, Nero had founded the quinquennial Neronia, a massive Greek-style festival in honor of his rule, in which he had competed publicly in poetry, music, and acting. Suetonius described Nero's renewed commitment to these performances after the fire. He expanded his artistic ambitions with a tour of Greece, where he appeared in numerous cities and forced judges to award him victory (Suetonius Nero 23). Suetonius made limited mention of Nero's relief efforts and reconstruction of Rome, focusing instead on his tours and artistic instincts in response to the disaster. He thus characterized Nero as a theatrical and incompetent emperor, implying a neglect of civic duty expected of the city's princeps.

Suetonius also described Nero's construction of the Domus Aurea as a violation of Roman norms of self-restraint. Built on land cleared by the fire, the palace had dining halls fitted with ivory ceilings and hidden machinery that released perfumes and flower petals onto guests. The grounds stretched across the city, filled with groves, meadows, and artificial lakes. Suetonius quoted Nero's reaction when the palace was finished: "now I can at last begin to live like a human being!" (31). This quote does not appear in other historical sources. Nevertheless, Suetonius placed it immediately after a detailed account of luxury, perhaps to emphasize the emperor's twisted priorities. In the same section, "Nero 31," Suetonius recorded that Nero planned to expand the palace's grounds all the way to Ostia, a seaport twenty-five kilometers from the center of Rome (31). This detail hinted that Nero's ambitions extended to a broader control of Roman territory. By placing the Domus Aurea in a negative light, Suetonius suggested that Nero sought to turn his city into a private imperial estate.

In conclusion, Suetonius's The Twelve Caesars focused primarily on the personal conduct of emperors, especially when it revealed moral failure or misconduct. The image of Nero as an arsonist survived because it fit the eccentric character that Suetonius had constructed in his biography of Nero. Unlike Tacitus, who examined the fire with relative objectivity, Suetonius's background as an imperial secretary allowed him to compile the sensational anecdotes that informed his largely satirical portrayal of the emperor. Thus, Suetonius is often regarded by modern historians as a biographer rather than a historian. His Life of Nero entertained his readers with vivid details, ultimately becoming one of the most enduring accounts of Nero's reign.

Cassius Dio's Roman History

In Roman History, Cassius Dio followed the course of Rome from its foundation to decline. Composed in the early third century CE (ca. 229 CE), his version remains one of the

most elaborate surviving accounts of the Great Fire, written more than a century and a half after the event. Compared to Tacitus or Suetonius, Dio was less concerned with firsthand accounts or direct access to imperial records, as he was a Greek-speaking senator under the Severan emperors writing in a period of political uncertainty. Nero and the fire became, in his telling, one of many examples of imperial corruption. Dio's depiction of Nero was influenced by Dio's senatorial position, his Greek background, and his aim to educate newer generations about the rise and fall of dynasties. It was written for readers long after the incident, meant to act as a lesson in how emperors could bring ruin through cruelty and pride.

Dio's portrait of the Great Fire was influenced by his position as a senator during the unstable Severan period. He spent much of his political career under emperors who held power through military support and hereditary succession. Soldiers elevated the general Septimius Severus to the throne after the civil war of 193 CE, making him the first emperor of the Severan dynasty (Herodian 2.9). Likewise, the emperor Caracalla's position as sole ruler was secured after the army backed his elimination of his brother, Geta (4.4). The Senate continued to meet, but its authority was limited, as key decisions were made primarily by the emperor and his advisors. Senators passed decrees, gave speeches, and took part in rituals, but these actions were often done under imperial commands. As a senator, Dio had experienced this firsthand. He described trials that had outcomes settled in advance, and he noted how accusations could be used to remove rivals. These purges and confiscations occurred across multiple reigns, such as those of Caracalla (198-217 CE) and Elagabalus (218-222 CE), and were often linked to changes in imperial favor. In Roman History, Dio mentioned that much was in turmoil in this period, demonstrating his skepticism toward the structures of imperial power under Severan emperors (Dio 72.36). In particular, he opposed the way authority was distributed and institutions functioned under military pressure. His concerns thus informed the way he documented earlier emperors from the Julio-Claudian dynasty.

Dio's Greek education and rhetorical technique influenced the structure and style of Roman History, including his account of the Great Fire. Raised in Bithynia in the Eastern Roman Empire, Dio wrote in Greek, which was the language of the intellectual elite in the East (Dio 80.5). He drew from the historiographical models of Greek historians Thucydides and Polybius rather than the Latin annalistic tradition. These Hellenistic histories often had a narrative arc, featuring tensions between two characters, reversals of fortune, and even visual scenes that resembled the tone of tragedy or theater. As such, Dio's history leaned primarily on dramatic scenes and presented events as part of a moral arc, including characters driven by ambition, arrogance and eventually destined for downfall. For example, in his account of Nero's early reign, Dio observed that the young emperor was "at first comparatively moderate" in his love for music and lyre-playing but later "began to indulge in each of these pursuits in a more open and precipitate fashion" (61.4). The portrait implied an unhealthy desire for music and performance from a young age, and the phrase "at first" foreshadows Nero's slow decline of discipline toward the middle years of his reign. Tacitus, writing in Latin, used fewer words. He described the same period as one of youthful extravagance and a fondness for the stage (Tacitus 14.14). Unlike

Tacitus's more straightforward and austere account, Dio tended to build vivid scenes and add color to his stories. As scholar Bill Thayer notes, Dio "steeped himself in the vocabulary and thoughts of [Thucydides'] great model until he could think almost unconsciously in the words of the other" (Thayer). By emphasizing detailed character arcs, Dio's Roman History resembled the dramatic pace of a Greek tragedy.

Dio's account of the Great Fire presented the event as a reflection of Nero's rule, both in its destruction and the emperor's response. The surviving account appears in Book 62 of Roman History. Although the original has been lost, the account was transcribed in an eleventh-century epitome compiled by John Xiphilinus, a monk and court scholar in Constantinople (Dio 62.16). His summary preserved the basic narrative but omitted much of Dio's wording and structure. It is thus difficult to recover Dio's tone from this version, since Xiphilinus compressed certain scenes and removed excess commentary. Even so, the preserved material presents a striking portrait of the emperor. While Tacitus only remarked on rumors of Nero singing about the fall of Troy, Dio—like Suetonius—claimed that Nero watched the flames from an elevated space, where "there was the best general view" of the fire, and performed songs about The Sack of Ilium (62.18). Similarly, Dio alluded to the Sibylline prophecies, a collection of ancient oracles consulted during Roman crises, noting that the fire prompted the populace to recall an omen declaring the "last of the sons of Aeneas, a mother-slayer shall govern" (62.18). Dio's inclusion of this prophecy implied that Nero would be the last of the imperial lineage, and the burning of Rome's foundations physically manifested the end of the Julio-Claudian line. Instead of detailing the fire's factual logistics, such as how it started or which districts were ruined, Dio tended to focus on the behavior and rumors surrounding the emperor. He suggested Nero's actions as signs of corruption and made his account of this moment relatively theatrical.

Dio further reported that men in disguise were seen spreading the fire, a claim likely based on a misreading of firefighting activity. In Book 62, he wrote that houses "were set on fire by the same men who came to lend assistance" and that "instead of putting out fires, [they] kindled new ones" (62.17). He did not identify these individuals directly, but implied they acted on imperial orders. The most likely explanation is that they were Vigiles, Rome's organized fire brigade, tasked with managing and extinguishing urban fires. The Vigiles were known to destroy buildings in the path of a blaze in order to create firebreaks, a practice that could easily be misunderstood in the chaos of a city-wide disaster. These efforts were most likely coordinated by Nero himself, who, despite his reputation, took an active role in organizing relief and response once the fire had spread. This interpretation is supported by several modern scholars. For instance, Anthony Barrett argues that Dio's agents were "almost certainly the emperor's appointees," and that their actions reflected "the perfectly normal process of creating firebreaks" (Barrett 129). This confusion may have resulted from Dio's temporal distance from the event, but it may also be possible that he repeated rumors without verifying them in a likely effort to reinforce Nero's image as a cruel ruler.

Dio's version of the fire differed distinctly from earlier accounts by Tacitus and Suetonius in tone, emphasis, and purpose. Like Tacitus, he blamed Nero, but he did so with more certainty

and less caution. Tacitus wrote that "whether [the fire was] accidental or treacherously contrived by the emperor, is uncertain," leaving space for ambiguity (Tacitus 15.38). Dio removed that uncertainty, and instead presented Nero's guilt as an established fact. Suetonius, too, described Nero's theatrical behavior during the fire, writing that he "watched the conflagration" and was "enraptured by what he called 'the beauty of the flames'" (Suetonius Nero 38). Dio included similar details, but integrated them into a structured account of Nero's collapse and corruption. In his version, Nero's actions were presented as consistent with the ongoing failures of his reign. This reflected Dio's aim to demonstrate the degeneration of Roman leadership, a theme he returned to throughout his later books.

Dio's Roman History was written with a distinct purpose to teach. He therefore presented the Great Fire as a fable, where the characters were clear, the consequences were harsh, and the moral was easy to grasp. To that end, he presented Nero as a villain whose actions, he suggested, brought harm to the city. To create a useful warning for readers, Dio wrote that Nero used the fire as a pretext to extract sums of money from citizens and cut off the public food supply. This line shows Nero exploiting a tragedy for personal gain, thereby encouraging readers to question Nero's sincerity (Dio 62.18). Such details gave Dio's story shape and direction, as they allowed readers to recognize Nero's perpetration of injustice, even if they lived centuries after the fire. Many of Dio's readers would never hold power themselves, but they might serve as senators or administrators under those who did. They would therefore need to judge character, recognize danger, and understand how power could corrupt. The fire's aftermath suggested imperial vanity and corruption to Roman society, and it could therefore act as an instructional example to reach many generations. Author T. D. Barnes notes that events within Roman History were often impacted by "[Dio's] own reflection and experience" (Barnes 253). That view explains why the fire appeared in a clear and moral frame in Dio's writing, since he used it as a lesson or warning.

To conclude, Cassius Dio's account of the Great Fire aimed to highlight Nero's deficiencies as a leader. His Roman History was intended to record events, but more importantly, to warn newer generations about the consequences of imperial power and vanity. Written from his perspective as a senator during the Severan dynasty, Dio's account reflected his critical views on the militaristic governance of the time. His Hellenistic and political influences contributed to an enduring, yet possibly exaggerated, image of Nero as a tyrant.

Conclusion

Many of the images that are now associated with Nero—the singer on the rooftop and the persecutor of Christians—come from the accounts of Tacitus, Suetonius, and Cassius Dio. Though vivid, these images were influenced largely by the authors' political contexts, audiences, and personal goals instead of by historical facts. Writing as senators or secretaries under different emperors, each historian used the Great Fire to express his anxieties about power and decline, while Nero's controversial personality helped ensure the sensational accounts would endure for two millennia. As the story passed from Tacitus to Suetonius to Dio, details like Nero singing

while Rome burned became increasingly elaborate and accusatory. By the time the written record settled, Nero had become the infamous arsonist of Rome.

Though each account differed in tone and emphasis, all three contributed to an identical portrayal of Nero as a corrupt and self-indulgent ruler. In the end, while Nero's actions played a role in tarnishing his legacy, the stories surrounding him persist to this day.

Works Cited

Barnes, T. D. "The Composition of Cassius Dio's 'Roman History.'" Phoenix, vol. 38, no. 3, 1984, pp. 240-55. JSTOR, https://doi.org/10.2307/1088277. Accessed 28 Aug. 2025.

Barrett, Anthony A. Rome Is Burning: Nero and the Fire That Ended a Dynasty. Princeton University Press, 2020.

Batiffol, Pierre. "L'église naissante: Introduction historique à l'étude du Nouveau Testament." Revue Biblique, vol. 3, 1864, pp. 514-15. Quoted in Anthony A. Barrett, Rome Is Burning: Nero and the Fire That Ended a Dynasty, Princeton UP, 2020, p. 167.

Brunt, P. A., and J. M. Moore, editors. Res Gestae Divi Augusti: The Achievements of the Divine Augustus. By Augustus, Oxford University Press, 1967.

Cretella, Frank Joseph. "Memory and Moderation: The Imperial Image of Augustus and Its Perception Among Later Authors." Master's thesis, The City University of New York, 2015. CUNY Academic Works, https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1906&context=gc_etds. Accessed 27 Aug. 2025.

Dando-Collins, Stephen. The Great Fire of Rome: The Fall of the Emperor Nero and His City. Da Capo Press, 2010.

Dio, Cassius. Roman History. Translated by Earnest Cary, Harvard University Press, 1914-1927. LacusCurtius: Into the Roman World, edited by Bill Thayer, University of Chicago, penelope.uchicago.edu/Thayer/E/Roman/Texts/Cassius_Dio/home.html. Accessed 5 Jul. 2025.

Eusebius of Caesarea. Church History. Translated by Arthur Cushman McGiffert, edited by Philip Schaff and Henry Wace, Buffalo, NY, Christian Literature Publishing Co., 1890. New Advent, revised and edited by Kevin Knight, 2026, www.newadvent.org/fathers/2501.htm. Accessed 10 Aug. 2025.

Flower, Harriet I. Roman Republics. Princeton University Press, 2010.

Griffin, Miriam T. Nero: The End of a Dynasty. Routledge, 1984.

Gualandi, Maria. Excavations on the North Slope of the Palatine. Rome Is Burning: Nero and the Fire That Ended a Dynasty, by Anthony A. Barrett, Princeton University Press, 2020, p. 75.

Herodian. Herodian's Roman History. Translated by Edward C. Echols, edited by Jona Lendering, Livius, 2020, www.livius.org/sources/content/herodian-s-roman-history/. Accessed 27 Aug. 2025.

Horace. Satires. Translated by A. S. Kline, Poetry in Translation, 2003, www.poetryintranslation.com/PITBR/Latin/HoraceSatiresBkISatI.php. Accessed 5 Jul. 2025.

Josephus, Flavius. The Antiquities of the Jews. Translated by William Whiston. Project Gutenberg, www.gutenberg.org/files/2848/2848-h/2848-h.htm. Accessed 10 Aug. 2025.

Juvenal. The Satires. Translated by A. S. Kline, Poetry in Translation, 2001, www.poetryintranslation.com/PITBR/Latin/JuvenalSatires3.php. Accessed 5 Jul. 2025.

Louis, André. Map of Ancient Rome. Rome Is Burning: Nero and the Fire That Ended a Dynasty, by Anthony A. Barrett, Princeton University Press, 2020, p. 15.

Pliny the Elder. The Natural History of Pliny. Translated by John Bostock and H. T. Riley, vol. 1, Henry G. Bohn, 1855. Internet Archive, archive.org/details/naturalhistoryof11855plin. Accessed 2 Aug. 2025.

Pseudo-Seneca. Octavia. Translated by Watson Bradshaw, London, 1902. Attalus, www.attalus.org/poetry/octavia.html. Accessed 10 Aug. 2025.

Romm, James. "Who Started It?" London Review of Books, vol. 43, no. 12, 17 June 2021, www.lrb.co.uk/the-paper/v43/n12/james-romm/who-started-it. Accessed 5 Jul. 2025.

Saumagne, Charles. “Les Incendiaires de Rome (ann. 64 p. C.) et les lois pénales des Romains (Tacite, Annales, XV, 44).” Revue Historique, vol. 227, 1962, pp. 337-60. Quoted in Anthony A. Barrett, Rome Is Burning: Nero and the Fire That Ended a Dynasty, Princeton UP, 2020, p. 171.

Seneca. De Clementia. Edited by Susanna Braund, Oxford University Press, 2009.

Severus, Sulpitius. The Sacred History of Sulpitius Severus. Translated by Alexander Roberts, edited by Philip Schaff and Henry Wace, Nicene and Post-Nicene Fathers, Second Series, vol. 11, Christian Classics Ethereal Library, 2006, www.ccel.org/ccel/schaff/npnf211.html. Accessed 10 Aug. 2025.

Skinner, Joseph. The Augustan Regiones. Map. Rome Is Burning: Nero and the Fire That Ended a Dynasty, by Anthony A. Barrett, Princeton University Press, 2020, p. 45.

Suetonius. The Lives of the Twelve Caesars. Translated by Alexander Thomson, revised by T. Forester, Project Gutenberg, 2004, www.gutenberg.org/files/6400/6400-h/6400-h.htm#link2H_4_0007. Accessed 1 Jul. 2025.

Tacitus. The Annals. Translated by Alfred John Church and William Jackson Brodribb, The Internet Classics Archive, Massachusetts Institute of Technology, classics.mit.edu/Tacitus/annals.11.xv.html. Accessed 28 Jun. 2025.

Tacitus. The Histories. Translated by Alfred John Church and William Jackson Brodribb, The Internet Classics Archive, Massachusetts Institute of Technology, classics.mit.edu/Tacitus/histories.1.i.html. Accessed 2 Aug. 2025.

Tacitus. The Life of Julius Agricola. Translated by A. S. Kline, 2015, Poetry in Translation, www.poetryintranslation.com/PITBR/Latin/TacitusAgricola.php. Accessed 2 Aug. 2025.

Thayer, Bill. Introduction. Roman History, by Cassius Dio, translated by Earnest Cary, Harvard UP, 1914. LacusCurtius, edited by Thayer, U of Chicago, 2021, penelope.uchicago.edu/Thayer/e/roman/texts/cassius_dio/introduction*.html. Accessed 27 Aug. 2025.

Vitruvius. De Architectura. Translated by Morris Hicky Morgan, Harvard University Press, 1914.

Yavetz, Zvi. “Forte an dolo principis (Tac. Ann. 15.38).” The Ancient Historian and His Materials: Essays in Honour of C. E. Stevens, edited by Barbara Levick, Gregg, 1975, pp. 181-97.

Melatonin Inhibits Human Leukemia U937 Cell Proliferation Via ERK Signaling Pathway Suppression By David Niu

Abstract

Objective: To investigate the effect of melatonin on the proliferation of U937 cells and to preliminarily explore its underlying molecular mechanisms. Methods: The Cell Counting Kit-8 (CCK-8) assay was used to detect changes in U937 cell viability after treatment with different concentrations of melatonin (0, 50, 100, 250, 500, 1000 μmol/L) for 24 h and 48 h. Cell cycle distribution and proliferation status were analyzed by PI staining and Ki-67 flow cytometry. The expression levels of proliferation-related proteins, including ERK and PCNA, were detected by Western blot. Results: Melatonin treatment significantly reduced U937 cell viability in a concentration-dependent manner, and the inhibitory effect was more pronounced at 48 h, showing a time-dependent tendency. Flow cytometry analysis showed that melatonin reduced the proportion of Ki-67-positive cells and induced cell cycle arrest at the G0/G1 and S phases. Western blot results revealed that treatment with 250 μmol/L melatonin significantly downregulated the expression of PCNA and ERK proteins. Conclusion: Melatonin can significantly inhibit the proliferation of U937 cells, and this effect may be mediated by inhibiting the ERK signaling pathway and downregulating PCNA expression.

Keywords: Melatonin; Leukemia; U937 cells; Cell proliferation; ERK; PCNA

Introduction

Leukemia is a malignant clonal hematopoietic disease originating from mutated hematopoietic stem and progenitor cells, characterized by massive accumulation of immature leukemic blasts in bone marrow and peripheral blood and subsequent impaired normal hematopoiesis (Bouligny et al.; Khoury et al.). Despite remarkable advances in conventional chemotherapy, targeted agents, and hematopoietic stem cell transplantation over the past decade, clinical obstacles including high recurrence rate, intrinsic or acquired chemoresistance, and severe systemic toxic complications persist in acute myeloid leukemia (AML) management (Wang et al.; Wu et al.). Under such a background, exploring low-toxic natural bioactive substances as auxiliary anti-leukemia candidates has become a hot spot of preclinical hematology research in recent years (Egbuna et al.; Hwang et al.).

Melatonin is an endogenous indole hormone primarily secreted by the pineal gland, participating in circadian rhythm modulation, antioxidant defense, and immune homeostasis maintenance (Reiter et al.; Socaciu et al.). Cumulative preclinical evidence over decades has validated the broad-spectrum anti-tumor property of melatonin across solid tumors and multiple hematologic malignancies via diverse signaling regulation (Mediavilla et al.; Talib et al.; Li et al.). In particular, numerous in vitro studies using HL-60, Jurkat, K562, and THP-1 myeloid leukemia cell lines have confirmed melatonin-mediated proliferation suppression and cell cycle disturbance (Lomovsky et al.; Quintana et al.; Ye et al.; Kasi et al.). Among the widely used

leukemia cell models for melatonin research, each cell line carries distinct lineage defects and experimental limitations: HL-60 is promyelocytic leukemia cells mainly applied for granulocytic differentiation research and lacks complete monocyte biological characteristics (Hornstein and Unfried; Basu et al.); K562 carries specific BCR-ABL fusion gene and good for chronic myeloid leukemia studies (Liu et al.); THP-1-derived macrophages are more responsive to M1 stimuli and skewed towards M1 phenotype (Nascimento et al.). In contrast, U937 is a classic pro-monocytic histiocytic leukemia cell line with a stable genetic background, unlimited in vitro subculture capacity, and consistent experimental repeatability. Notably, U937 cells inherently express functional MT1 melatonin receptors (RADOGNA et al.), which have been linked to melatonin-mediated regulation of monocytic leukemia cell proliferation and ERK cascade signaling (Chen et al.). Therefore, this study selects U937 as the core experimental cell model to carry out relevant verification. Recent reviews have highlighted the pleiotropic anti-leukemic effects of melatonin, with the ERK/MAPK pathway emerging as one of the potentially important signaling cascades (Mafi et al.; Moloudizargari et al.)

ERK cascade is the core downstream effector of MAPK signaling, governing physiological proliferation and G1/S phase transition; constitutive hyperphosphorylation of ERK is frequently detected in primary AML samples and cultured leukemic cells, which sequentially elevates Cyclin D1 and PCNA expression to drive unlimited malignant DNA replication (Pearson et al.; Chambard et al.; Steelman et al.). As an indispensable cofactor of DNA polymerase, PCNA expression level closely correlates with tumor proliferative activity and is widely used as a proliferation biomarker (Moldovan et al.; Mailand et al.; Kasprzak). U937 human monocytic leukemia cell endogenously expresses functional MT1 melatonin receptor, which provides the molecular prerequisite for melatonin's direct biological action on this cell model (RADOGNA et al.). Existing data have confirmed that melatonin inhibits ERK activity in THP-1 cells, whereas most previous investigations only described single-phase cell cycle arrest in various leukemic strains and lacked systematic clarification on the complete ERK-Cyclin D1-PCNA signaling cascade in U937 (X.-W. Liu et al.; Hsieh et al.; Kulsoom et al.). Moreover, standardized IC_{50} and sublethal concentration data of melatonin for U937 have not been documented in existing literature (Yoo and Jeung; Quintana et al.).

Based on the above unresolved research gaps, the current study adopted CCK-8, Ki67-PI flow cytometry, and western blot to systematically assess melatonin’s anti-proliferative effect on U937 cells and decode its underlying ERK-dependent molecular mechanism, aiming to supplement existing research data and lay an experimental foundation for follow-up preclinical transformation of melatonin in monocytic leukemia adjuvant treatment.

Materials and Methods

Materials

U937 cells were purchased from ATCC. RPMI‑1640 medium, fetal bovine serum (FBS), melatonin, CCK‑8 kit, RIPA lysis buffer, 10% PAGE gel rapid preparation kit, BCA protein

assay kit, ECL chemiluminescence reagent, and PI kit were purchased from Fisher Scientific. Main antibodies, including PCNA antibody, ERK antibody, Vinculin antibody, and HRP-conjugated secondary antibodies, were purchased from Fisher Scientific.

Cell culture

U937 cells were purchased from a certified cell resource bank. Cells were cultured in RPMI-1640 complete medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Cells were maintained in a humidified incubator at 37°C with 5% CO_2 and subcultured every 2–3 days to maintain logarithmic growth. All experiments were performed using cells with good growth status and confirmed free of contamination.

Melatonin treatment

Melatonin was purchased from MCE and dissolved in dimethyl sulfoxide (DMSO) to prepare a 100 mM stock solution, which was stored at −20°C protected from light. Before experiments, the stock solution was serially diluted with a complete medium to obtain working solutions at different final concentrations (0, 50, 100, 250, 500, 1000 μmol/L). The final DMSO concentration was kept consistent across all experimental groups to exclude solvent interference. After seeding, cells were allowed to stabilize, then treated with different concentrations of melatonin for the indicated times before subsequent assays.

CCK-8 assay for cell viability

U937 cells in logarithmic growth phase were collected and counted, then seeded into 96-well plates at a density of approximately 1×10^4 cells/well, with 3–5 replicate wells per group. After cells were stably suspended, different concentrations of melatonin were added for treatment for 24 h or 48 h. After treatment, 10 μL of CCK-8 reagent was added to each well according to the manufacturer's instructions, mixed gently, and incubated at 37°C for 1–2 h in the dark. Absorbance at 450 nm (OD value) was measured using a microplate reader, and relative cell viability was calculated based on the control group. The experiment was repeated at least three times to ensure reliability.

Flow cytometric detection of Ki-67 expression

Cells were seeded in 6-well plates and treated with 250 μmol/L melatonin for 24 h and 48 h, then harvested. After fixation and permeabilization, cells were incubated with a fluorescence-labeled anti-Ki-67 antibody in the dark. After washing, the percentage of Ki-67-positive cells was detected by flow cytometry.

PI staining for cell cycle

After treatment with 250 μmol/L melatonin for 24 h and 48 h, cells were collected and fixed in pre-cold 70% ethanol at 4°C overnight. After removing the fixative, cells were stained with propidium iodide (PI) staining solution containing RNase A and incubated for 30 min in the

dark. Data were acquired by flow cytometry, and the proportions of cells in G0/G1, S, and G2/M phases were calculated using analysis software.

Western Blot for protein expression

Treated cells were collected, washed twice with pre-chilled PBS, and lysed with an appropriate amount of RIPA lysis buffer (containing protease and phosphatase inhibitors) on ice for 30 min with intermittent vortexing. The lysate was centrifuged at 12,000 rpm for 15 min at 4°C, and the supernatant was collected as total protein. Protein concentration was determined using the BCA method, and equal amounts of protein from each group were separated by SDS-PAGE and transferred to PVDF membranes. Membranes were blocked with 5% non-fat milk or BSA for 1 h at room temperature, then incubated with primary antibodies overnight at 4°C. The next day, membranes were washed three times with TBST (5 min each) and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature. After washing again, protein bands were visualized using ECL chemiluminescence reagent and images were captured with a gel imaging system. Grayscale values of protein bands were quantified using image analysis software, and normalized to the internal control protein.

Statistical analysis

Data are presented as mean ± SD. One-way analysis of variance (ANOVA) and Student's t-test were used for statistical comparison. $P<0.05$ was considered statistically significant. * represents $P<0.05$.

Results

Melatonin inhibits U937 cell proliferation

As shown in Figure 1A and 1B, the CCK-8 assay demonstrated that U937 cell viability was reduced after treatment with different concentrations of melatonin compared to the control group. As the concentration of melatonin increased from 50 μmol/L to 1000 μmol/L, cell viability gradually decreased, showing a clear concentration-dependent difference ($P<0.05$). As shown in Figure A, at 24 h of treatment, moderate- and high-concentration groups already exhibited a certain degree of inhibition; as shown in Figure B, at 48 h of treatment, cell viability in all concentration groups was further reduced, indicating that the inhibitory effect of melatonin on U937 cells is time-dependent. As shown in Figure C, the CCK-8 results demonstrated that melatonin inhibited U937 cell viability in a dose-dependent manner, with an IC50 of 939.1 μmol/L. Considering that 250 μmol/L melatonin already exerted a certain effect on cell proliferation while maintaining cell viability at a relatively high level, thereby effectively avoiding non-specific cytotoxicity induced by high drug concentrations, 250 μmol/L was selected for subsequent experiments.

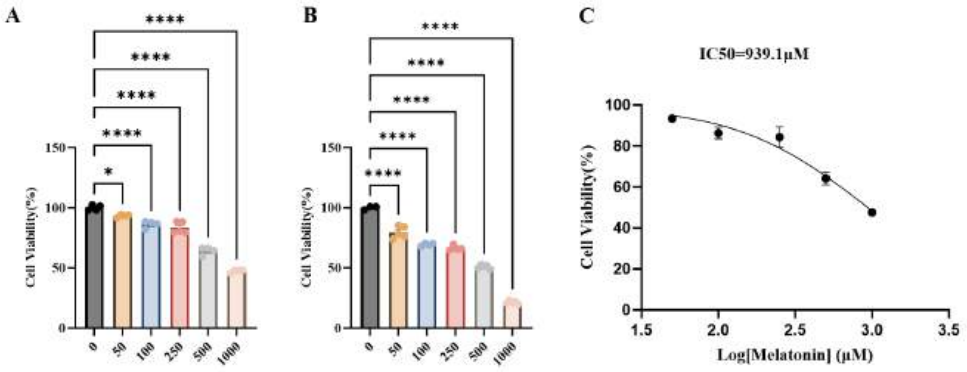


Figure 1: Melatonin inhibits the proliferation viability of U937 cells. (A-B) CCK-8 assay was used to detect cell viability after treatment with different concentrations of melatonin (0, 50, 100, 250, 500, 1000 μmol/L) for 24 h and 48 h. (C) Nonlinear regression fitting curve and half-maximal inhibitory concentration (IC50=939.1 μM) after melatonin treatment. Data are presented as mean ± SD (n=3). One-way ANOVA was used for statistical comparison. *p<0.05, ****p<0.0001 vs. control group (0 μmol/L).

Melatonin significantly reduces the expression of the proliferation marker Ki-67 in U937 cells

As shown in Figure 2A and 2B, Ki-67 is a nuclear protein expressed during all active phases of the cell cycle (G1, S, G2, M) and is a classic marker for assessing cell proliferation activity. Flow cytometry analysis showed that after treatment with 250 μmol/L melatonin for 24 h, the Ki-67-positive rate in U937 cells was significantly lower than that in the control group (P<0.05). When the treatment time was extended to 48 h, the Ki-67-positive rate in the melatonin group decreased further and sharply (P<0.01). This result directly demonstrates that melatonin induces a large number of U937 cells to exit the active proliferative state.

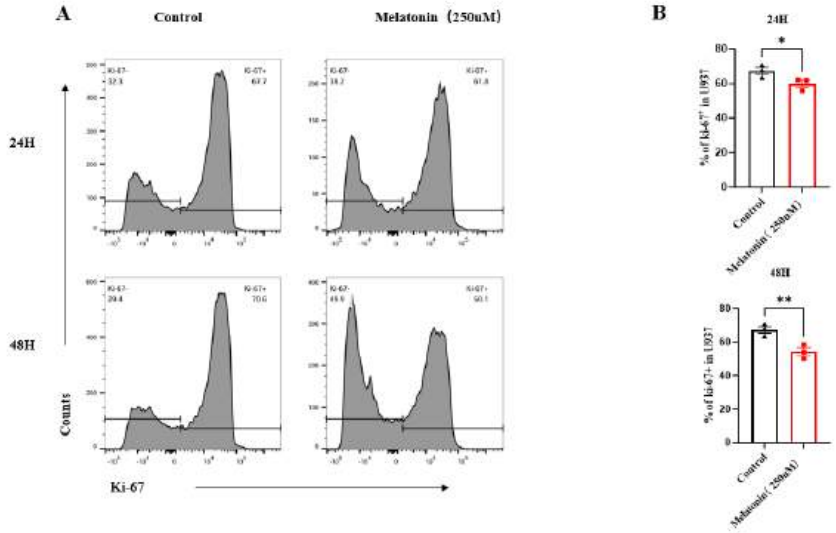


Figure 2: Melatonin inhibits cell proliferation. (A) Flow cytometry detection of Ki-67-positive cell ratios in control and melatonin (250 μmol/L) treated groups at 24 h and 48 h; the right panel shows the quantitative bar graph. (B) Quantitative bar graph for flow cytometry analysis. Data are presented as mean ± SD (n=3). An independent samples t-test was used for statistical comparison. *p<0.05, **p<0.01, ***p<0.001. ns, not significant.

Melatonin induces G0/G1 and S phase arrest in U937 cells

As shown in Figure 3A and 3B, to further clarify the specific nodes of proliferation inhibition, PI staining and flow cytometry were performed for cell cycle distribution analysis. The results showed that treatment with 250 μmol/L melatonin for 24 h caused a significant increase in the proportion of S-phase cells. At 48 h after treatment, cell cycle redistribution occurred: compared with the control group, the proportion of cells in G0/G1 phase was significantly increased in the melatonin-treated group ($P<0.001$), the proportion in S phase was also significantly increased ($P<0.001$), accompanied by a very significant decrease in the proportion of cells in G2/M phase ($P<0.0001$). This "dual arrest" phenomenon suggests that melatonin not only prevents cells from transitioning from G1 to S phase but may also interfere with the DNA synthesis process during S phase, preventing cells from completing replication and entering G2/M phase.

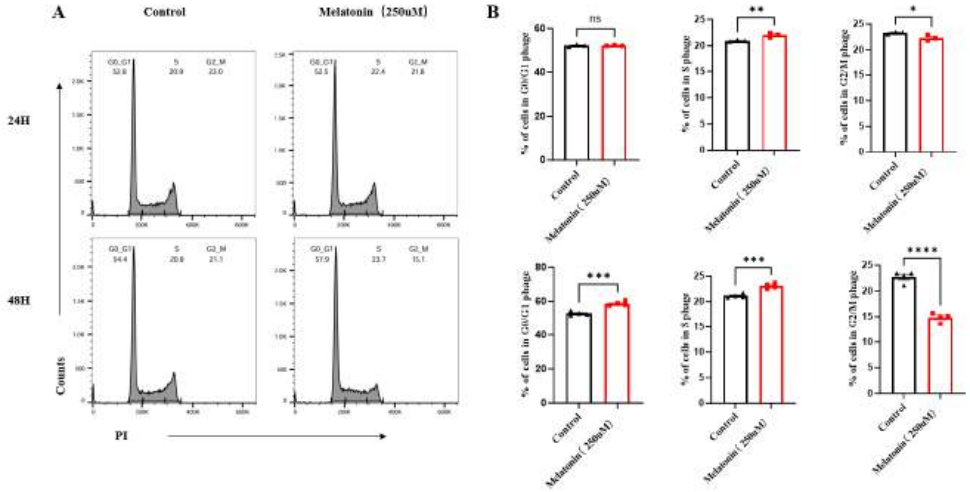


Figure 3 Melatonin induces cell cycle arrest. (A) Cell cycle distribution detected by PI staining combined with flow cytometry after treatment for 24 and 48 hours; (B) quantitative analysis of the percentages of cells in G0/G1, S, and G2/M phases in the right panel. Data are presented as Mean ± SD (n=3). Statistical differences were compared using an independent samples t-test, *p<0.05, p<0.01, ***p<0.001, ****p<0.0001. ns indicates no significant difference.

Melatonin downregulates Cyclin-D1 and PCNA expression by blocking the ERK pathway

As shown in Figure 4A, given the cell cycle arrest indicated by flow cytometry, we further explored the underlying molecular drivers using Western blot. The MAPK/ERK signaling pathway is a hub regulating cell proliferation and cell cycle progression. The results showed that after treatment with 250 μmol/L melatonin for 48 h, total ERK protein expression was downregulated, and its phosphorylation level (p-ERK) was strongly inhibited. Further examination of downstream targets revealed that the expression of Cyclin-D1, a key protein driving cells through the G1/S checkpoint, decreased sharply. In addition, the expression of PCNA, a classic protein associated with DNA synthesis, was also significantly downregulated, which is highly consistent with the observed decrease in Ki-67 levels and S-phase arrest.

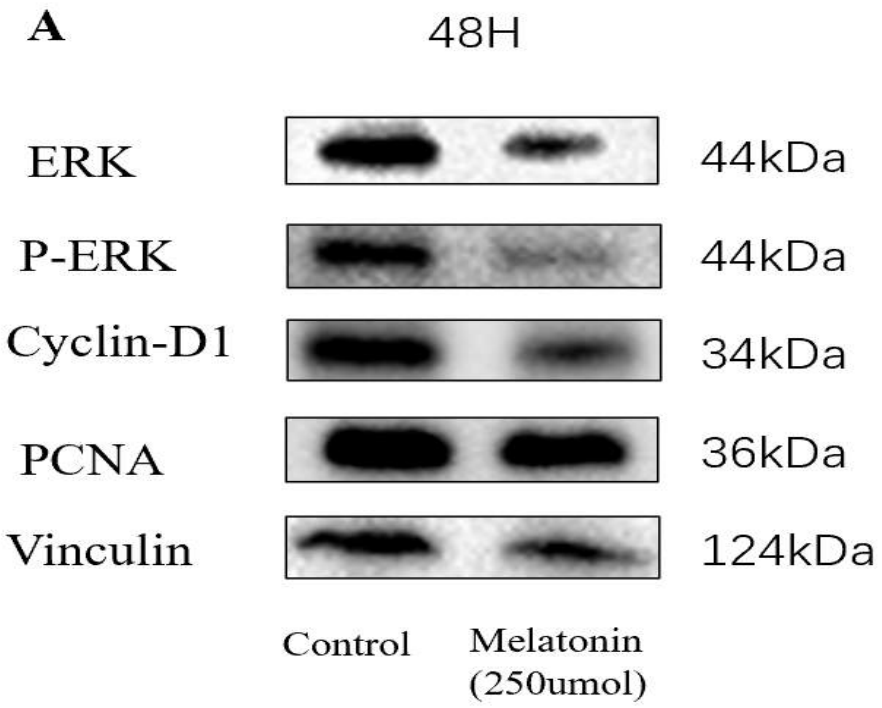


Figure 4: Melatonin regulates the expression of proteins related to cell proliferation and the ERK signaling pathway. (A) Western blot detection of total ERK, phosphorylated ERK (p-ERK1/2), Cyclin-D1, and proliferating cell nuclear antigen (PCNA) protein expression levels in control and melatonin (250 μmol/L) treated groups for 48 h. Vinculin was used as an internal control to ensure equal loading.

Discussion

The progression of hematological malignancies is tightly associated with aberrant cell cycle regulation and persistent overactivation of mitogenic signaling pathways; therefore, the identification of low-toxicity natural anti-proliferative agents has attracted considerable research interest. As an endogenous hormone with superior biosafety, melatonin's anti-neoplastic potential has been extensively validated in abundant preclinical research (Mediavilla et al.; T. Li et al.). Combined phenotypic data from CCK-8, flow cytometry, and western blot, the present work elaborates melatonin's inhibitory function on U937 proliferation and its downstream molecular regulation route, and further compares our findings with previously published data to clarify the supplementary value of our research.

Our CCK-8 quantitative data revealed melatonin suppresses U937 viability in a concentration- and time-dependent manner, with calculated IC_{50} values of 939.1 μM and 250 μM screened as non-cytotoxic sublethal working concentrations for subsequent functional detection. While most prior studies exploring melatonin on U937 arbitrarily set drug dosage without IC_{50} quantification and sublethal concentration screening, our quantified numerical data therefore supplement blank basic dosage parameters for subsequent in vivo experimental design. Under 250μM treatment, the percentage of Ki-67-positive proliferative U937 cells was markedly decreased, which is consistent with the previous research outcomes in breast carcinoma models

that melatonin restrains tumor proliferation via reducing Ki67 expression (Jardim-Perassi et al.; Sonehara et al.), further confirming melatonin works mainly through cytostasis rather than non-specific cytotoxicity.

Cell cycle detection via PI staining showed melatonin treatment for 48 h simultaneously elevated the proportion of U937 cells at G0/G1 and S phases while reducing the G2/M population, which differs substantially from previously reported cell cycle phenotypes in related cancer research. Available literature commonly indicated melatonin triggers isolated G1 arrest in HL-60, Jurkat, and MCF-7 cells, and exclusive G2/M arrest in glioma cell lines (Rubio et al.; Gil et al.). Even a recent in vitro study on THP-1 only observed single-cycle retardation after melatonin intervention. From a cell cycle kinetic perspective, elevated G0/G1 percentage indicates G1/S checkpoint blockage caused by lowered Cyclin D1, while accumulated S-phase cells result from impaired DNA replication due to PCNA downregulation (Zhou et al.; Lomovsky et al.).

Molecular mechanism exploration revealed that inactivation of the ERK signaling pathway and the consequent cascade inhibition of its downstream target proteins are central to driving the above cell cycle arrest. The MAPK/ERK signaling cascade plays a key role in transmitting mitogenic signals and driving cell cycle progression. The present study found that melatonin significantly inhibited ERK phosphorylation (p-ERK) activation in U937 cells. As a result of this decreased upstream kinase activity, the expression of Cyclin-D1, a core cell cycle regulatory protein that drives cells from G1 to S phase, was drastically reduced, which at the molecular level explains why a large number of cells fail to pass the G1/S checkpoint and remain arrested in the G0/G1 phase. Simultaneously, the protein expression of PCNA, a core auxiliary factor for DNA polymerase, was also significantly downregulated. The deficiency of PCNA implies a defect in the assembly and function of DNA replication during S phase, which is highly consistent with the S-phase retardation and decreased Ki-67 expression observed in flow cytometry. Existing solid tumor-related publications have repeatedly confirmed melatonin-induced ERK deactivation (Mediavilla et al.; T. Li et al.), but relevant leukemia research was mostly fragmented: scattered reports only detected partial ERK inhibition without continuous observation of downstream Cyclin D1 and PCNA changes. Our study establishes an intact ERK-pERK-Cyclin D1-PCNA axis regulatory cascade in U937, linking upstream signaling inhibition to dual-cycle arrest, which makes up for the incomplete pathway evidence in previous related studies, further highlighting the translational value of our molecular findings for subsequent melatonin formulation optimization.

Although this study elucidated the anti-proliferative mechanism of melatonin at the cellular and molecular levels, certain limitations remain. For example, whether sustained cell cycle arrest under longer-term, high-dose drug treatment would further trigger a caspase-dependent apoptotic program requires verification in subsequent experiments. Moreover, the mechanistic findings based on in vitro cell models need to be further evaluated using animal models to assess the in vivo anti-leukemia efficacy of melatonin. Therefore, follow-up work will construct a U937 xenograft mouse model to translate our in vitro results into in vivo preclinical

data and clarify receptor-dependent regulatory mechanisms.

Conclusion

In summary, melatonin inhibits ERK phosphorylation in U937 monocytic leukemia cells, sequentially reduces Cyclin D1 and PCNA expression, and consequently induces unique concurrent G0/G1 and S-phase arrest to retard malignant proliferation. Distinguished from former relevant reports, our study supplements three meaningful research outcomes: we quantified the specific IC_{50} and sublethal working concentration of melatonin against U937; identified an unprecedented dual G0/G1+S cell cycle arrest phenotype in melatonin-treated leukemic cells; and constructed a complete ERK-Cyclin D1-PCNA regulatory cascade to interpret its anti-proliferative mechanism. Our findings indicate that melatonin holds potential for leukemia therapy, and further studies, including animal experiments, are warranted to validate its anti-leukemic efficacy

Works Cited

Basu, Jhinuk, et al. “Molecular and Epigenetic Alterations in Normal and Malignant Myelopoiesis in Human Leukemia 60 (HL60) Promyelocytic Cell Line Model.” Frontiers in Cell and Developmental Biology, vol. 11, 2023, doi:10.3389/fcell.2023.1060537.

Bouligny, Ian M., et al. “Mechanisms of Myeloid Leukemogenesis: Current Perspectives and Therapeutic Objectives.” Blood Reviews, vol. 57, 2023, p. 100996, doi:10.1016/j.blre.2022.100996.

Chambard, Jean-Claude, et al. “ERK Implication in Cell Cycle Regulation.” Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, vol. 1773, no. 8, 2007, pp. 1299–310, doi:10.1016/j.bbamcr.2006.11.010.

Chen, Min, et al. “Melatonin MT_1 and MT_2 Receptor ERK Signaling Is Differentially Dependent on $G_{i/o}$ and $G_{q/11}$ Proteins.” Journal of Pineal Research, vol. 68, no. 4, 2020, doi:10.1111/jpi.12641.

Egbuna, Chukwuebuka, et al. “Phytochemicals and Bioactive Compounds Effective against Acute Myeloid Leukemia: A Systematic Review.” Food Science & Nutrition, vol. 11, no. 7, 2023, pp. 4191–210, doi:10.1002/fsn3.3420.

Gil, Beatriz Irene Fernandez, et al. “Melatonin Disrupts Glioblastoma Metabolism and Enhances Temozolomide Cytotoxic Effects.” Neurosurgery, vol. 66, no. Supplement 1, 2019, pp. 310–644, doi:10.1093/neuros/nyz310_644.

Hornstein, Tamara, and Klaus Unfried. “Differentiation of HL-60 Cells into Primed Neutrophils for the Evaluation of Antiapoptotic Effects of Poorly Soluble Nanoparticles.” PLOS One, vol. 20, no. 7, 2025, p. e0328717, doi:10.1371/journal.pone.0328717.

Hsieh, Tzuo-Yi, et al. “Melatonin Induces Cell Cycle Arrest and Suppresses Tumor Invasion in Urinary Bladder Urothelial Carcinoma.” Aging, 2023, doi:10.18632/aging.204673.

Hwang, Dongwon, et al. “Natural Products and Acute Myeloid Leukemia: A Review Highlighting Mechanisms of Action.” Nutrients, 2019, doi:10.3390/nu11051010.

Jardim-Perassi, Bruna Victorasso, et al. “Effect of Melatonin on Tumor Growth and Angiogenesis in the Xenograft Model of Breast Cancer.” PLoS ONE, vol. 9, no. 1, 2014, p. e85311, doi:10.1371/journal.pone.0085311.

Kasi, Reena, et al. “Melatonin Induces Apoptosis and Inhibits the Proliferation of Cancer Cells via Reactive Oxygen Species-Mediated MAPK and MTOR Pathways.” Clinical Cancer Drugs, vol. 7, no. 1, 2020, pp. 44–56, doi:10.2174/2212697x06666191116151114.

Kasprzak, Aldona. “Prognostic Biomarkers of Cell Proliferation in Colorectal Cancer (CRC): From Immunohistochemistry to Molecular Biology Techniques.” Cancers, vol. 15, no. 18, 2023, p. 4570, doi:10.3390/cancers15184570.

Khoury, Joseph D., et al. “The 5th Edition of the World Health Organization Classification of Haematolymphoid Tumours Myeloid and Histiocytic Dendritic Neoplasms.” Leukemia, vol. 36, 2022, pp. 1703–19, doi:10.1038/s41375-022-01613-1.

Kulsoom, Kulsoom, et al. “Revealing Melatonin’s Mysteries: Receptors, Signaling

Pathways, and Therapeutic Applications." Hormone and Metabolic Research, vol. 56, no. 06, 2023, pp. 405–18, doi:10.1055/a-2226-3971.

Li, Li, et al. "Loss of Keratin 23 Enhances Growth Inhibitory Effect of Melatonin in Gastric Cancer." Molecular Medicine Reports, vol. 29, no. 2, 2023, doi:10.3892/mmr.2023.13145.

Li, Tian, et al. "Melatonin: Does It Have Utility in the Treatment of Haematological Neoplasms?" British Journal of Pharmacology, vol. 175, no. 16, 2017, pp. 3251–62, doi:10.1111/bph.13966.

Li, Ya, et al. "Melatonin for the Prevention and Treatment of Cancer." Oncotarget, vol. 8, no. 24, 2017, pp. 39896–921, doi:10.18632/oncotarget.16379.

Liu, Xin-Wei, et al. "Melatonin Exerts Protective Effect on N2a Cells under Hypoxia Conditions through Zip1/ERK Pathway." Neuroscience Letters, vol. 595, 2015, pp. 74–80, doi:10.1016/j.neulet.2015.04.013.

Liu, Yan-Hong, et al. "ND-09 Inhibits Chronic Myeloid Leukemia K562 Cell Growth by Regulating BCR-ABL Signaling." Oncology Reports, vol. 46, no. 1, 2021, doi:10.3892/or.2021.8087.

Lomovsky, A. I., et al. "Melatonin Enhances the Effect of ABT-737 in Acute Monocytic Leukemia THP-1 Cells." Молекулярная Биология, vol. 58, no. 1, 2024, pp. 141–53, doi:10.31857/s0026898424010136.

Mafi, Alireza, et al. "Melatonin as a Regulator of Apoptosis in Leukaemia: Molecular Mechanism and Therapeutic Perspectives." Frontiers in Pharmacology, vol. 14, 2023, doi:10.3389/fphar.2023.1224151.

Mailand, Niels, et al. "Regulation of PCNA–Protein Interactions for Genome Stability." Nature Reviews Molecular Cell Biology, vol. 14, no. 5, 2013, pp. 269–82, doi:10.1038/nrm3562.

Mediavilla, M. D., et al. "Basic Mechanisms Involved in the Anti-Cancer Effects of Melatonin." Current Medicinal Chemistry, vol. 17, no. 36, 2010, pp. 4462–81, doi:10.2174/092986710794183015.

Moldovan, George-Lucian, et al. "PCNA, the Maestro of the Replication Fork." Cell, vol. 129, no. 4, 2007, pp. 665–79, doi:10.1016/j.cell.2007.05.003.

Moloudizargari, Milad, et al. "Therapeutic Targets of Cancer Drugs: Modulation by Melatonin." Life Sciences, vol. 267, 2021, p. 118934, doi:10.1016/j.lfs.2020.118934.

Nascimento, Camyla Rodrigues, et al. "Comparison of Monocytic Cell Lines U937 and THP-1 as Macrophage Models for in Vitro Studies." Biochemistry and Biophysics Reports, vol. 32, 2022, p. 101383, doi:10.1016/j.bbrep.2022.101383.

Pearson, Gray, et al. "Mitogen-Activated Protein (MAP) Kinase Pathways: Regulation and Physiological Functions*." Endocrine Reviews, vol. 22, no. 2, 2001, pp. 153–83, doi:10.1210/edrv.22.2.0428.

Quintana, Carlos, et al. "Melatonin Enhances Hyperthermia‐induced Apoptotic Cell Death in Human Leukemia Cells." Journal of Pineal Research, vol. 61, no. 3, 2016, pp. 381–95,

doi:10.1111/jpi.12356.

RADOGNA, FLAVIA, et al. “Melatonin as an Apoptosis Antagonist.” Annals of the New York Academy of Sciences, vol. 1090, no. 1, 2006, pp. 226–33, doi:10.1196/annals.1378.025.

Reiter, Russel J., et al. “Melatonin as a Mitochondria-Targeted Antioxidant One of Evolution’s Best Ideas.” Cell Mol Life Sci, 2017, doi:10.1007/s00018-017-2609-7.

Rubio, Sara, et al. “Inhibition of Proliferation and Induction of Apoptosis by Melatonin in Human Myeloid HL‐60 Cells.” Journal of Pineal Research, vol. 42, no. 2, 2007, pp. 131–38, doi:10.1111/j.1600-079x.2006.00392.x.

Samanta, Saptadip. “Melatonin: An Endogenous Miraculous Indolamine, Fights against Cancer Progression.” Journal of Cancer Research and Clinical Oncology, vol. 146, no. 8, 2020, pp. 1893–922, doi:10.1007/s00432-020-03292-w.

Socaciu, Andreea Iulia, et al. “Melatonin, an Ubiquitous Metabolic Regulator: Functions, Mechanisms and Effects on Circadian Disruption and Degenerative Diseases.” Reviews in Endocrine and Metabolic Disorders, vol. 21, no. 4, 2020, pp. 465–78, doi:10.1007/s11154-020-09570-9.

Sonehara, Nath&#xfffd;lia, et al. “Melatonin Regulates Tumor Aggressiveness under Acidosis Condition in Breast Cancer Cell Lines.” Oncology Letters, 2018, doi:10.3892/ol.2018.9758.

Steelman, L. S., et al. “Roles of the Ras/Raf/MEK/ERK Pathway in Leukemia Therapy.” Leukemia, vol. 25, no. 7, 2011, pp. 1080–94, doi:10.1038/leu.2011.66.

Talib, Wamidh H., et al. “Melatonin in Cancer Treatment: Current Knowledge and Future Opportunities.” Molecules, vol. 26, no. 9, 2021, p. 2506, doi:10.3390/molecules26092506.

Wang, Bofei, et al. “Therapeutic Hurdles in Acute Myeloid Leukemia: Leukemic Stem Cells, Inflammation and Immune Dysfunction.” Current Opinion in Pharmacology, vol. 82, 2025, p. 102526, doi:10.1016/j.coph.2025.102526.

Wu, Xia, et al. “Advances in Drug Delivery Systems for the Treatment of Acute Myeloid Leukemia.” Small, 2024, doi:10.1002/smll.202403409 .

Ye, Kaiqin, et al. “Melatonin Sensitizes Leukemia Cells to the MCL1 Inhibitors S63845 and A‐1210477 through Multiple Pathways.” Journal of Pineal Research, vol. 76, no. 2, 2024, doi:10.1111/jpi.12943.

Yoo, Yeong-Min, and Eui-Bae Jeung. “Melatonin-Induced Calbindin-D9k Expression Reduces Hydrogen Peroxide-Mediated Cell Death in Rat Pituitary GH3 Cells.” Journal of Pineal Research, vol. 48, no. 2, 2010, pp. 83–93, doi:10.1111/j.1600-079x.2009.00730.x.

Zhou, Mengxi, et al. Melatonin Induces Cell Cycle Arrest, Inhibits Cell Proliferation and Accelerates Apoptosis by Modulation of CDK4 in NSCLC. 2022, doi:10.21203/rs.3.rs-1230442/v1.

Basu, Jhinuk, et al. “Molecular and Epigenetic Alterations in Normal and Malignant Myelopoiesis Human Leukemia 60 (HL60) Promyelocytic Cell Line Model.” Frontiers in Cell and Deve Biology, vol. 11, 2023, doi:10.3389/fcell.2023.1060537.

Bouligny, Ian M., et al. “Mechanisms of Myeloid Leukemogenesis: Current Perspectives and The Objectives.” Blood Reviews, vol. 57, 2023, p. 100996, doi:10.1016/j.blre.2022.100996.

Chambard, Jean-Claude, et al. “ERK Implication in Cell Cycle Regulation.” Biochimica et Biophy (BBA) - Molecular Cell Research, vol. 1773, no. 8, 2007, pp. 1299–310, doi:10.1016/j.bbamcr.2006.11.010.

Chen, Min, et al. “Melatonin MT_1 and MT_2 Receptor ERK Signaling Is Differentially Dependent and $G_{q/11}$ Proteins.” Journal of Pineal Research, vol. 68, no. 4, 2020, doi:10.1111/jpi.12641

Egbuna, Chukwuebuka, et al. “Phytochemicals and Bioactive Compounds Effective against Acute Leukemia: A Systematic Review.” Food Science & Nutrition, vol. 11, no. 7, 2023, pp. 419 doi:10.1002/fsn3.3420.

Gil, Beatriz Irene Fernandez, et al. “Melatonin Disrupts Glioblastoma Metabolism and Enhances Temozolomide Cytotoxic Effects.” Neurosurgery, vol. 66, no. Supplement 1, 2019, pp. 31(doi:10.1093/neuros/nyz310_644.

Hornstein, Tamara, and Klaus Unfried. “Differentiation of HL-60 Cells into Primed Neutrophils f Evaluation of Antiapoptotic Effects of Poorly Soluble Nanoparticles.” PLOS One, vol. 20, 2025, p. e0328717, doi:10.1371/journal.pone.0328717.

Hsieh, Tzuo-Yi, et al. “Melatonin Induces Cell Cycle Arrest and Suppresses Tumor Invasion in U Bladder Urothelial Carcinoma.” Aging, 2023, doi:10.18632/aging.204673.

Hwang, Dongwon, et al. “Natural Products and Acute Myeloid Leukemia: A Review Highlighting Mechanisms of Action.” Nutrients, 2019, doi:10.3390/nu11051010.

Jardim-Perassi, Bruna Victorasso, et al. “Effect of Melatonin on Tumor Growth and Angiogenesis in Xenograft Model of Breast Cancer.” PLoS ONE, vol. 9, no. 1, 2014, p. e85311, doi:10.1371/journal.pone.0085311.

Kasi, Reena, et al. “Melatonin Induces Apoptosis and Inhibits the Proliferation of Cancer Cells via Reactive Oxygen Species-Mediated MAPK and MTOR Pathways.” Clinical Cancer Drugs, vol. 7, no. 1, 2020, pp. 44–56, doi:10.2174/2212697x06666191116151114.

Kasprzak, Aldona. “Prognostic Biomarkers of Cell Proliferation in Colorectal Cancer (CRC): From Immunohistochemistry to Molecular Biology Techniques.” Cancers, vol. 15, no. 18, 2023, p. 4570, doi:10.3390/cancers15184570.

Khoury, Joseph D., et al. “The 5th Edition of the World Health Organization Classification of Haematolymphoid Tumours Myeloid and Histiocytic Dendritic Neoplasms.” Leukemia, vol. 36, 2022, pp. 1703–19, doi:10.1038/s41375-022-01613-1.

Kulsoom, Kulsoom, et al. “Revealing Melatonin’s Mysteries: Receptors, Signaling Pathways, and Therapeutic Applications.” Hormone and Metabolic Research, vol. 56, no. 06, 2023, pp. 405–18, doi:10.1055/a-2226-3971.

Li, Li, et al. “Loss of Keratin 23 Enhances Growth Inhibitory Effect of Melatonin in Gastric

Cancer." Molecular Medicine Reports, vol. 29, no. 2, 2023, doi:10.3892/mmr.2023.13145.

Li, Tian, et al. "Melatonin: Does It Have Utility in the Treatment of Haematological Neoplasms?" British Journal of Pharmacology, vol. 175, no. 16, 2017, pp. 3251–62, doi:10.1111/bph.13966.

Li, Ya, et al. "Melatonin for the Prevention and Treatment of Cancer." Oncotarget, vol. 8, no. 24, 2017, pp. 39896–921, doi:10.18632/oncotarget.16379.

Liu, Xin-Wei, et al. "Melatonin Exerts Protective Effect on N2a Cells under Hypoxia Conditions through Zip1/ERK Pathway." Neuroscience Letters, vol. 595, 2015, pp. 74–80, doi:10.1016/j.neulet.2015.04.013.

Liu, Yan-Hong, et al. "ND-09 Inhibits Chronic Myeloid Leukemia K562 Cell Growth by Regulating BCR-ABL Signaling." Oncology Reports, vol. 46, no. 1, 2021, doi:10.3892/or.2021.8087.

Lomovsky, A. I., et al. "Melatonin Enhances the Effect of ABT-737 in Acute Monocytic Leukemia THP-1 Cells." Молекулярная Биология, vol. 58, no. 1, 2024, pp. 141–53, doi:10.31857/s0026898424010136.

Mafi, Alireza, et al. "Melatonin as a Regulator of Apoptosis in Leukaemia: Molecular Mechanism and Therapeutic Perspectives." Frontiers in Pharmacology, vol. 14, 2023, doi:10.3389/fphar.2023.1224151.

Mailand, Niels, et al. "Regulation of PCNA–Protein Interactions for Genome Stability." Nature Reviews Molecular Cell Biology, vol. 14, no. 5, 2013, pp. 269–82, doi:10.1038/nrm3562.

Mediavilla, M. D., et al. "Basic Mechanisms Involved in the Anti-Cancer Effects of Melatonin." Current Medicinal Chemistry, vol. 17, no. 36, 2010, pp. 4462–81, doi:10.2174/092986710794183015.

Moldovan, George-Lucian, et al. "PCNA, the Maestro of the Replication Fork." Cell, vol. 129, no. 4, 2007, pp. 665–79, doi:10.1016/j.cell.2007.05.003.

Moloudizargari, Milad, et al. "Therapeutic Targets of Cancer Drugs: Modulation by Melatonin." Life Sciences, vol. 267, 2021, p. 118934, doi:10.1016/j.lfs.2020.118934.

Nascimento, Camyla Rodrigues, et al. "Comparison of Monocytic Cell Lines U937 and THP-1 as Macrophage Models for in Vitro Studies." Biochemistry and Biophysics Reports, vol. 32, 2022, p. 101383, doi:10.1016/j.bbrep.2022.101383.

Pearson, Gray, et al. "Mitogen-Activated Protein (MAP) Kinase Pathways: Regulation and Physiological Functions*." Endocrine Reviews, vol. 22, no. 2, 2001, pp. 153–83, doi:10.1210/edrv.22.2.0428.

Quintana, Carlos, et al. "Melatonin Enhances Hyperthermia‐induced Apoptotic Cell Death in Human Leukemia Cells." Journal of Pineal Research, vol. 61, no. 3, 2016, pp. 381–95, doi:10.1111/jpi.12356.

RADOGNA, FLAVIA, et al. "Melatonin as an Apoptosis Antagonist." Annals of the New York Academy of Sciences, vol. 1090, no. 1, 2006, pp. 226–33, doi:10.1196/annals.1378.025.

Reiter, Russel J., et al. "Melatonin as a Mitochondria-Targeted Antioxidant One of Evolution's Best Ideas." Cell Mol Life Sci, 2017, doi:10.1007/s00018-017-2609-7.

Rubio, Sara, et al. "Inhibition of Proliferation and Induction of Apoptosis by Melatonin in Human Myeloid HL‐60 Cells." Journal of Pineal Research, vol. 42, no. 2, 2007, pp. 131–38, doi:10.1111/j.1600-079x.2006.00392.x.

Samanta, Saptadip. “Melatonin: An Endogenous Miraculous Indolamine, Fights against Cancer Progression.” Journal of Cancer Research and Clinical Oncology, vol. 146, no. 8, 2020, pp. 1893–922, doi:10.1007/s00432-020-03292-w.

Socaciu, Andreea Iulia, et al. “Melatonin, an Ubiquitous Metabolic Regulator: Functions, Mechanisms and Effects on Circadian Disruption and Degenerative Diseases.” Reviews in Endocrine and Metabolic Disorders, vol. 21, no. 4, 2020, pp. 465–78, doi:10.1007/s11154-020-09570-9.

Sonehara, Nath&#xfffd;lia, et al. “Melatonin Regulates Tumor Aggressiveness under Acidosis Condition in Breast Cancer Cell Lines.” Oncology Letters, 2018, doi:10.3892/ol.2018.9758.

Steelman, L. S., et al. “Roles of the Ras/Raf/MEK/ERK Pathway in Leukemia Therapy.” Leukemia, vol. 25, no. 7, 2011, pp. 1080–94, doi:10.1038/leu.2011.66.

Talib, Wamidh H., et al. “Melatonin in Cancer Treatment: Current Knowledge and Future Opportunities.” Molecules, vol. 26, no. 9, 2021, p. 2506, doi:10.3390/molecules26092506.

Wang, Bofei, et al. “Therapeutic Hurdles in Acute Myeloid Leukemia: Leukemic Stem Cells, Inflammation and Immune Dysfunction.” Current Opinion in Pharmacology, vol. 82, 2025, p. 102526, doi:10.1016/j.coph.2025.102526.

Wu, Xia, et al. “Advances in Drug Delivery Systems for the Treatment of Acute Myeloid Leukemia.” Small, 2024, doi:10.1002/smll.202403409 .

Ye, Kaiqin, et al. “Melatonin Sensitizes Leukemia Cells to the MCL1 Inhibitors S63845 and A‐1210477 through Multiple Pathways.” Journal of Pineal Research, vol. 76, no. 2, 2024, doi:10.1111/jpi.12943.

Yoo, Yeong-Min, and Eui-Bae Jeung. “Melatonin-Induced Calbindin-D9k Expression Reduces Hydrogen Peroxide-Mediated Cell Death in Rat Pituitary GH3 Cells.” Journal of Pineal Research, vol. 48, no. 2, 2010, pp. 83–93, doi:10.1111/j.1600-079x.2009.00730.x.

Zhou, Mengxi, et al. Melatonin Induces Cell Cycle Arrest, Inhibits Cell Proliferation and Accelerates Apoptosis by Modulation of CDK4 in NSCLC. 2022, doi:10.21203/rs.3.rs-1230442/v1.

The Role of Financial Influencers and Social Media Streamers in Executing Pump-and-Dump Schemes Across Memecoin and Meme Stock Markets By Dev Kalaria

Abstract

How do celebrities, financial influencers, and social media streamers drive pump-and-dump schemes in memecoin markets? This paper investigates market manipulations to show how modern promotion systematically extracts capital from retail investors. By analyzing three different case studies (Dogecoin, SaveTheKids Token, and Hawk Tuah coin), this study demonstrates how different actors exploit specific target audiences. The findings reveal that retail losses do not stem from ordinary market volatility or random chance. Instead, they are driven by an asymmetric information architecture designed to ensure everyday traders serve as exit liquidity for insiders. Data shows that while influencer promotions generate a brief 1.83% return on day zero, they trigger a 6.53% cumulative loss by day 30, compounding to a massive 62.8% annualized loss for holders. This exploitation relies on modern tactics that legacy disclosure laws cannot effectively regulate, including ambient celebrity memes, hidden token ownership, and pre-trade supply manipulation. Because current investor protections are built for traditional equity markets with clear issuers and cash payments, they fail to adequately regulate decentralized token dynamics entirely. Ultimately, protecting retail capital requires a complete shift away from outdated stock market rules and toward frameworks built specifically for decentralized network structures.

1. Introduction

Celebrities and influencers shape cryptocurrency markets in ways that systematically transfer wealth from retail investors to insiders. This paper examines how these individuals influence cryptocurrency markets and who ultimately bears the financial burden. The rise of celebrity-driven cryptocurrency promotion has exposed a structural gap in financial consumer protection: social media platforms now function as unregulated distribution channels for investment influence. Current regulatory debates have converged on a pointed question: do celebrities and influencers who dominate social media pose a material threat to retail investors in the memecoin market? The cryptocurrency market peaked at $3 trillion in 2021, but prices now move more on influencer posts than fundamentals. High-profile figures and organizations like Elon Musk, FaZe Clan, and Donald Trump have used their platforms to promote various tokens, often leading to rapid, speculative price surges. These surges are frequently followed by precipitous declines that inflict significant losses on late-arriving retail investors.

A single social media post can send a token's price soaring within hours. Ante (2023), using an event study approach across 47 cryptocurrency-related Twitter events, found significant positive abnormal returns and trading volume following Elon Musk's posts, with price effects proving especially significant for Dogecoin related tweets specifically. Merkley et al. (2024) analyzed nearly 36,000 tweets from 180 prominent crypto social media accounts and found that while influencer mentions generate an average initial return of 1.83% on the day of posting,

cumulative average returns fall to approximately -6.53% by day 30. This pattern is consistent with deliberate pump-and-dump dynamics: retail investors purchase assets driven by speculative momentum and absorb subsequent depreciation, while promoters who liquidate their positions early secure substantial profits. This paper examines the mechanisms that make this transfer of wealth possible and the structural conditions that enable it.

2 Analytical Framework

This paper's framework analyzes how asymmetric information architectures enable project organizers to manipulate token supply and timing prior to public promotion. It incorporates narrative economics and parasocial interaction theory to evaluate how influencers convert audience trust into immediate, coordinated retail buying pressure. Finally, the model examines regulatory arbitrage to identify how decentralized asset mechanics bypass legacy disclosure laws through hidden token ownership.

3. Background

3.1 Meme Coin Market

Unlike traditional equities or fixed-income securities, memecoins lack fundamental valuation metrics such as cash flow or revenue. Instead, they rely entirely on social sentiment and collective belief. In fact, research shows that fewer than 2% of memecoins survive in the long term [6]. Because these markets are driven by everyday retail investors rather than institutions, prices move purely on speculation and social media trends. Ultimately, narrative control directly translates to price control.

The structural vulnerability is compounded by market size. La Morgia et al. (2023) found that 83.5% of tokens targeted in coordinated pump-and-dump schemes had less than $1 million in daily trading volume, and 71% had market capitalizations below $20 million [1].

Characteristic	**Share of targeted coins**
Daily trading volume under $1 million	83.5%
Market capitalization below $20 million	71%
Priced below $0.40	78.3%
Market capitalization below $50 million	Most frequently targeted

Table 1: Structural Characteristics of Targeted Pump-and-Dump Tokens
Source [1]

Research analyzing pump-and-dump frequency and effectiveness establishes that these schemes are most frequent and impactful in cryptocurrencies with a market capitalization below $50 million, because these low-capitalization coins require relatively little capital to manipulate [3]. A coordinated purchase, amplified by social media coverage, can generate artificial price momentum that influencers then broadcast to their audiences as evidence of organic growth.

3.2 The Rise of Influencer Promotion

Influencers actively engineer market sentiment to exploit information asymmetry, creating a structural disadvantage that consistently results in capital losses for retail investors. While peer-reviewed literature highlights a clear trend of growing influence across Twitter, Reddit, and TikTok, the underlying mechanism driving these losses is structural inequality. Influencers, unlike retail investors, control timing, narrative, and asset visibility, allowing them to accumulate positions before broadcasting to their audience. By the time a retail investor sees the post, the influencer or their associates have typically already accumulated a position. For instance, the Hawk Tuah Coin case illustrates this extreme asymmetry; 10 insider-controlled developer wallets captured 96% of the token supply at launch. Although the project team publicly blamed "sniper bots" for the lopsided distribution as a damage control tactic, the centralization ensured that public promotion served only as an exit liquidity event for the organizers.

This form of structural exploitation renders traditional financial analysis ineffective. Academic studies demonstrate that crypto influencers rely on emotion-driven content rather than fundamentals, yet their messaging dictates market prices. Using Shiller's concept of narrative economics, influencers do not simply reflect market sentiment; they actively shape it in ways that drive behavior independent of asset value. Even without formal promotion (such as Elon Musk's indirect influence through Dogecoin memes), their large audiences create the same conditions that enable manipulation. As a result, retail investors remain at a structural disadvantage, unable to match the informational and timing advantages of promoters.

3.3 Three Key Groups

Market influence in the memecoin sector operates through three structurally distinct categories of public figures, each with a different relationship to their audience, a different mechanism of credibility, and a different speed of effect. Merkley et al. (2024) note that crypto influencers operate similarly to coordinated pump organizers, often acting as information gatekeepers who trigger initial price movements, and that their structural position controlling timing, narrative, and asset visibility allows them to shape investor behavior at scale [2]. Understanding how each group leverages its position is essential before analyzing the specific investor segments they target.

4 Methods and Material

This study employs a multi-case research design evaluating Dogecoin, SaveTheKids Token, and Hawk Tuah Coin to identify distinct promoter dynamics. The methodology utilizes quantitative analysis, incorporating event study methodologies that track price velocity, abnormal trading volumes, and cumulative returns across specific promotional windows. Furthermore, data collection relies on cross-platform social media text scraping coupled with on-chain ledger forensics to map developer wallet distributions. Finally, the empirical framework incorporates behavioral finance metrics and gambling psychology data to measure retail sentiment shifts and community transaction frequencies.

5 Discussion

5.1 Influencing Retail Investors

Retail investors occupy a structurally disadvantaged position that ensures they are the final participants in the information chain. This vulnerability is the core principle of modern pump-and-dump architectures, not an incidental side effect. The lifecycle of these schemes begins well before any public signal exists, as organizers pre-select and purchase target tokens while deliberately withholding the asset's identity to preserve their absolute informational advantage. In the Hawk Tuah Coin case, this architecture reached its extreme when 10 insider and connected developer wallets captured 96% of the token supply at the moment of launch. The project team subsequently attributed this concentration to external sniper bots to divert attention from internal manipulation, preventing retail buyers from participating in genuine price discovery entirely.

After this stealth accumulation phase, organizers enter a coordinated build-up window designed to engineer mass psychology. Livshits and Zhao (2021) document this chronological sequence in detail by analyzing coordinated pump events across various cryptocurrency exchanges [11]. They show that administrators publicly announce a specific date, time, and exchange for an upcoming pump event while keeping the target coin strictly anonymous. They then flood their channels with repeated, high-frequency reminders in the hours leading up to the signal to maximize audience anticipation and fear of missing out (FOMO). When the coin is finally revealed, a massive wave of retail market orders triggers an instantaneous trading volume and price spike. At the peak of this artificial surge, organizers and insiders systematically liquidate their holdings before everyday investors can process the shift.

The velocity of this reversal is exceptionally rapid. La Morgia et al. (2023) found that standard pump-and-dump operations last only a few minutes from genesis to collapse, with their detection model successfully identifying active pumps within a mere 25 seconds [1]. This incredibly tight window demonstrates that retail investors are structurally unable to exit positions before the crash. Because decentralized markets rely on immediate liquidity pools, the massive, automated sell-offs by insiders instantly drain the available buy orders. Consequently, retail traders attempting to sell face massive price slippage or transaction failures, while the organizers' high-priority scripts outbid standard retail transactions for network processing.

This structural delay is further compounded by advanced market signals. Later research by Livshits and Zhao (2021) demonstrated that unusual transaction volume or sudden price movements in the hours leading up to a public reveal are far more effective predictors of the target coin than the actual social announcements themselves [11]. By analyzing these hidden on-chain patterns, the target token can be identified within a narrow 6.1% error margin before the organizers even broadcast its name to their channels. This proves that a secondary layer of insider buying occurs well before the official launch sequence. Ultimately, every crucial variable that determines the financial outcome of the trade, including asset accumulation, narrative build-up, and liquidation timing, has already concluded before a single retail investor encounters the promotional signal.

Time period after influencer mention	**Average return**
Day 0 (day of mention)	+1.83%
Day 2 to Day 5	−1.02%
Day 10	−2.24%
Day 30	−6.53%
Annualized loss (30-day holders)	−62.8%

Source [11]

5.2 Celebrity Influence

Celebrity endorsements often replace traditional due diligence in memecoin markets, making highly speculative assets appear culturally credible. Token organizers routinely leverage the visibility of public figures to maximize audience reach while obscuring the underlying financial mechanics. The Hawk Tuah Coin launch exemplifies this dynamic perfectly. Organizers constructed a financial vehicle directly around Hailey Welch’s peak viral relevance to mask a profit ceiling established by insider wallets. Whether Welch was an active participant, an uninformed promoter, or an exploited figurehead remains undetermined; however, the final outcome for retail buyers is exactly the same. The damage caused by this information gap operates independently of the promoter's actual intent.

Elon Musk’s involvement with Dogecoin ($DOGE) illustrates a related concept known as ambient influence. By deploying casual humor, memes, and indirect social media posts instead of formal financial disclosures, Musk shifted massive market capitalizations without triggering standard advertising regulations. Ante (2023) highlights these complex market dynamics, noting that Tesla made a $1.5 billion corporate acquisition of Bitcoin during periods of heightened social media commentary by Musk [4]. This introduced clear conflicts of interest that remained completely hidden from retail followers.

Across both cases, the operational reality is clear. Project creators do not select public figures for their financial literacy. Instead, they choose them purely for their cultural reach and audience trust, leaving a major regulatory gap regarding accountability for non-traditional financial promotion.

5.3 Influencing Existing Communities

Pre-existing online communities are structurally vulnerable to manipulation by token organizers because the very features that enable efficient information sharing also enable coordinated exploitation. This vulnerability does not arise from inexperience, but paradoxically from familiarity with crypto markets. First, many participants have prior trading experience and understand volatility and pump-and-dump dynamics, which leads them to falsely believe they can time entry and exit effectively. Second, this overconfidence reduces due diligence, as early price movements are interpreted as confirmation of individual judgment rather than as evidence of manipulation. Third, these communities already possess shared language, mutual trust, and established decision-making norms, allowing organizers to mobilize coordinated buying behavior without building new networks. Consequently, prior experience and community cohesion do not protect participants; they increase susceptibility to manipulation.

5.4 How Financial Influencers Co-opt Existing Communities

Financial influencers exploit existing communities by monetizing established trust to mobilize coordinated buying. Horton and Wohl (1956) defined parasocial interaction as a one-sided emotional bond that audiences develop with media figures who remain unaware of their existence [12]. In crypto communities, this effect is amplified because the influencer is not viewed as an outsider; rather, they are someone the community has followed, trusted, and traded alongside for a long time. When that influencer promotes a token, members often perceive it as confirmation from an insider rather than as advertising or manipulation. Consequently, the line between genuine advice and controlled promotion becomes difficult to discern.

Community members do not act irrationally when they follow these signals; they rely on previously trusted patterns that are now artificially engineered. Kim et al. (2023) demonstrated this in their analysis of r/WallStreetBets during the GameStop short squeeze: a 272% increase in Reddit comments predicted a 17% rise in trading turnover within 30 minutes, with even larger effects in options markets [15]. Critically, the researchers found no connection between that comment activity and abnormal stock returns. The trading was noise, not informed action. Participants were not deceived by ignorance; they responded to deliberately manufactured social signals that structurally mimicked legitimate ones.

Furthermore, participants in many of these trading communities exhibit a heightened vulnerability to manipulation. Research on gambling psychology indicates that 57.3% of surveyed crypto investors were considered at-risk gamblers, while 15.6% were classified as

problem gamblers [13]. This study demonstrated that higher gambling risk correlates directly with stronger FOMO and impulsivity [13].

Risk classification	Surveyed Crypto Investors
At-risk gamblers	57.3%
Problem gamblers	15.6%
Non-risk gamblers	27.1%

Source [13]

Even experienced traders remain vulnerable to emotional decisions. In many crypto communities, the most active and vocal members often exhibit these tendencies. Their excitement projects an illusion of community confidence, while their impulsive behavior causes them to buy quickly after an influencer's signal and hold longer after the dump begins. Although their experience may help them react faster, their underlying psychology makes them easier to exploit.

Pump-and-dump schemes are designed to make every participant feel like an insider. La Morgia et al. (2023) document how these operations use a ranking system: higher-ranked members receive the coin signal earlier and buy at lower prices, while lower-ranked members buy in later as the price rises [1]. Ranks are earned by recruiting new members, which expands the buyer pool below the recruiter. Every tier believes their position gives them an edge over the one below, yet none can see the organizers sitting above the entire structure, already ready to sell. Experienced traders are not immune to this dynamic; they compete to climb the hierarchy, recruiting the next layer of buyers while unknowingly serving as exit liquidity for the layer above them.

5.5 How Streamers Create Trust

Social identity and community belonging actively override individual risk assessment in real-time streaming environments. When a streamer displays a live price chart during a session, the desire to remain aligned with group consensus creates social pressure that neutralizes protective instincts. The SaveTheKids Token, promoted by FaZe Clan members, exploited the tight-knit subscriber communities they had built over years. The live streaming environment made questioning the token's supply distribution socially expensive and technically difficult, preventing experienced members from exercising the due diligence they might apply in isolated environments.

This highlights the structural difference between streamer influence and ambient celebrity influence, such as what Musk exercised over Dogecoin. Musk's tweets operated on a group of isolated individuals each making a private decision, whereas a streamer operates on a community making a visible, collective decision. The social pressure in the latter case is intentional. A

viewer who sells while the stream is live, or who types skepticism into the chat, does so against the consensus of hundreds of other participants. The experienced trader's instincts, which are meant to protect them, are neutralized by the social cost of acting against the community's visible momentum.

5.6 The Limitations of Regulation in the Kardashian case

The SEC's October 2022 case against Kim Kardashian highlights both the power and the strict limits of current crypto regulations. Kardashian paid $1.26 million in penalties, disgorgement, and interest to settle charges that she promoted EthereumMax tokens on Instagram without disclosing her $250,000 she received to promote the coin. She agreed to a three-year crypto promotion ban, reinforcing SEC Chair Gary Gensler's warning that celebrities must disclose the source and amount of any payment received for promoting crypto securities [14]. However, this precedent, the most aggressive regulatory action to date, also proves what existing frameworks cannot reach, as the SEC could only act because EthereumMax met the legal definition of a security and Kardashian received a documented payment she failed to disclose. Most influencer dynamics that harm retail investors bypass these exact conditions: Elon Musk’s Dogecoin posts involved no documented payment and an asset that is not legally a security; the FaZe Clan influencers in the SaveTheKids case held token ownership rather than receiving a payment, exploiting a loophole the laws do not clearly cover; and the Hawk Tuah Coin organizers relied on bot-driven supply capture during a token launch rather than a misleading promotional statement. Ultimately, retail investors remain vulnerable because the mechanisms driving their losses fall entirely outside the scope of the Kardashian precedent.

5.7 Remaining Gaps and Future Directions

Machine learning research has focused on detecting pump-and-dump events in real time before retail investors act on the signal. A 2024 paper from the University of Zurich presents a real-time prediction pipeline that uses natural language processing to monitor Telegram channels, classifying messages into the different phases of a pump-and-dump cycle and identifying target coins before the signal is publicly released [16]. In a live test on the Poloniex exchange, the model accurately predicted the target coin among the top five candidates in 24 out of 43 events, demonstrating that these schemes leave detectable patterns even before the signal is broadcast [16]. A separate system called Perseus, deployed in real conditions between February and October 2024, uses graph networks to trace the organizers behind pump-and-dump activity across social media platforms, successfully identifying 438 scheme masterminds during its deployment [18]. However, the limitation is structural: these systems were built to catch coordinated Telegram-based schemes of the kind La Morgia et al. document [1], not the decentralized narrative dynamics through which DOGE and Hawk Tuah Coin operated. A model trained on Telegram message patterns cannot flag a celebrity's tweet as a pump signal because the tweet looks identical to ordinary cultural commentary.

Another possible solution is increased financial literacy. Research on gambling psychology found that cryptocurrency literacy was one of the few variables that meaningfully reduced FOMO-driven investment decisions among at-risk investors [13]. Reddit communities like r/CryptoCurrency have developed informal self-protective heuristics, such as checking token supply distribution on-chain, verifying team wallets, and looking for code modifications like the anti-whale alteration in the SaveTheKids case. Although these practices are not peer-reviewed or formalized, they show that informed community members can identify the same red flags that cost retail investors millions in the cases examined in this paper.

The gap this paper identifies is not that solutions do not exist. It is that the solutions available were built for a simpler version of the problem than the one harming retail investors at scale.

6 Conclusion

The structural failure of memecoin markets stems from an information architecture designed to ensure retail investors arrive last. As the trajectories of Dogecoin, SaveTheKids, and Hawk Tuah Coin demonstrate, modern token manipulation purposely bypasses traditional consumer protections. Existing disclosure laws, highlighted by the SEC's 2022 Kardashian precedent, only work if three conditions are met: there is a documented cash payment, the asset legally qualifies as a security, and the promoter fails to disclose the payout.

Memecoin schemes are built to avoid these exact triggers. Subtle shifts in ambient social media commentary involve no direct marketing payouts, influencer compensation is built directly into hidden token ownership, and automated insider wallets capture the supply before public trading even begins. Because these tactics operate outside the reach of legacy stock market rules, protecting everyday investors requires a complete regulatory evolution. Policymakers must move past outdated disclosure laws and design new frameworks built specifically for the realities of decentralized promotion, automated supply manipulation, and real-time social media influence.

Works Cited

Ante, L. “How Elon Musk's Twitter Activity Moves Cryptocurrency Markets.” Technological Forecasting and Social Change, vol. 186, 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3778844

Ardia, D., and Bluteau, K. “Twitter and Cryptocurrency Pump-and-Dumps.” International Review of Financial Analysis, vol. 95, 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4456159

Bolz, Manuel, et al. “Machine Learning-Based Detection of Pump-and-Dump Schemes in Real-Time.” arXiv, 2024. https://arxiv.org/abs/2412.18848

Griffiths, Mark D., et al. “Investigating the Role of Regret, FOMO and Financial Literacy in Cryptocurrency Speculation.” International Journal of Mental Health and Addiction, Springer, 2025. https://link.springer.com/article/10.1007/s11469-025-01555-6

Horton, Donald, and R. Richard Wohl. “Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance.” Psychiatry, vol. 19, no. 3, 1956, pp. 215–229.

Jagolinzer, Alan D. “Market and Regulatory Implications of Social Identity Cohorts: A Discussion of Crypto Influencers.” Review of Accounting Studies, vol. 29, 2024, pp. 2298–2317. https://doi.org/10.1007/s11142-024-09837-5

Kim, Kyunghee, et al. “Social Informedness and Investor Sentiment in the GameStop Short Squeeze.” Electronic Markets, vol. 33, no. 1, 2023, article 23. https://doi.org/10.1007/s12525-023-00632-9

La Morgia, M., Mei, A., Sassi, F., and Stefa, J. “The Doge of Wall Street: Analysis and Detection of Pump and Dump Cryptocurrency Manipulations.” ACM Transactions on Internet Technology, vol. 23, no. 1, 2023. https://arxiv.org/pdf/2105.00733

Livshits, Victor, and Jiahua Zhao. “Crypto Pump-and-Dump Schemes: Detection and Prediction via Behavioral Signals.” Expert Systems with Applications, vol. 182, 2021, article 115284. https://www.sciencedirect.com/science/article/abs/pii/S0957417421007156

Mancino, Davide. “The Memecoin Phenomenon: An In-Depth Study of Solana's Blockchain Trends.” arXiv, 2025. https://arxiv.org/abs/2512.11850

Merkley, Kenneth J., et al. “Crypto-Influencers.” Review of Accounting Studies, vol. 29, 2024, pp. 2254–2297. https://doi.org/10.1007/s11142-024-09838-4

Meyer, Eva Andrea, et al. “Testing the Credibility of Crypto Influencers: An Event Study on Bitcoin.” Finance Research Letters, vol. 60, 2024, article 104864. https://doi.org/10.1016/j.frl.2023.104864

Mondal, Lubdhak, et al. “Causality between Sentiment and Cryptocurrency Prices.” arXiv, 2023. https://arxiv.org/abs/2306.05803

“Pump and Dump Schemes in Cryptocurrency Markets.” IEEE Xplore, 2019. https://ieeexplore.ieee.org/abstract/document/8955594

Sundarasen, Sheela, and Farida Saleem. “From Tweets to Trades: A Bibliometric and Systematic Review of Social Media's Influence on Cryptocurrency.” International Journal of Financial Studies, vol. 13, no. 2, 2025, p. 87. https://doi.org/10.3390/ijfs13020087

United States Securities and Exchange Commission. “SEC Charges Kim Kardashian for Unlawfully Touting Crypto Security.” SEC.gov, 3 Oct. 2022, www.sec.gov/newsroom/press-releases/2022-183.

The Effect of Demographics on Organic Food Perception By Venkata Anish Neti

Abstract

Organic food is more influential and controversial than ever before with the organic market growing every year. However, a declining number of people actually know what the organic label stands for and how it affects a consumers' perception. Current research has addressed this idea, however, very few studies have looked into the effect of demographics on organic food perception (specifically value and price perception, and their relationship). Adults between the ages of 18-80 are the ones currently buying food and who this research would affect the most. Participants took a survey designed to assess their overall value of the organic label and their overall perception of the price of food based on the label, both in relation with their demographic groups, all to determine the correlation between demographic group and perception. A multitude of statistical analysis strategies determined that price perception is a more general cognitive bias, while value perception is more individualistic based on demographic, and that price and value perception acted independently. Future research is required to find an actual possible causal relationship between demographics and perception. To do this an actual experiment must be conducted with randomly selected participants.

Keywords Organic Food, Perception, Price, Value, Demographics, Willingness to Pay

Introduction

"Don't judge a book by its cover" is a phrase everyone has heard; it's often used to teach children not to accept things at face value and to question things. Yet as we get older, many of us seem to forget this lesson—especially concerning our food. Most adults who purchase organic products do not even consider what the organic label actually signifies or what standards it represents. That said, over the past few decades, organic food has become increasingly popular and controversial. In just the last twenty years, the organic market has expanded exponentially, generating billions of dollars and creating millions of jobs. Despite this growth, there remains confusion about what the organic label actually guarantees. According to previous studies, many consumers misunderstand organic labels, often believing that "organic" means "healthier" or "higher quality," even though the certification is only about production methods, not nutritional value (Hansen, 2013; Cohrssen & Miller, 2016). Furthermore, while there has been research conducted on general attitudes toward organic products, few investigations have examined how consumers from various demographic groups view and justify the higher prices associated with organic products. This gap is important because understanding how demographics shape perception and spending can improve food labeling policies and consumer education. Therefore, this study seeks to answer the question: To what extent does the presence of the organic label on food products influence the overall perception of the product by consumers of various demographic groups?

Definitions

Organic Food	- Products that have been produced using cultural, biological, and mechanical practices that cycle farm resources, create ecological balance, and improve biodiversity (McEvoy, 2016).
FOPNL	- Front of Package Nutrition Label (Penzavecchia et al., 2022, pp. 3037-3051).
Halo Effect	- A cognitive bias in which one positive attribute leads consumers to assume other unrelated positive qualities (Andreoni, 1990).
Warm Glow Effect	- A psychological phenomenon where individuals experience emotional satisfaction from performing actions they perceive as socially or environmentally good (Andreoni, 1990).
WTP (Willingness to Pay)	- The maximum price a consumer is willing to pay for a product (Bredar, 2015).
Price Premium	- The additional cost consumers are willing to pay for a labeled product compared to a similar non-labeled product (Rao & Monroe, 1989, pp. 351-357).

History of the Organic Label

When debating the organic label and its effects, an often brought up issue is the meaning and trustworthiness of the organic label. The organic label's creation and regulation is extremely complex, but in the simplest terms, FOPNL systems, like the organic label, are generally implemented to communicate complex nutritional information in a way that consumers can understand, to guide, inform, and shape consumer choices and behavior (Kanter et al., 2018, pp. 1399-1408). Out of the general category of organic labels, the most specific is the USDA organic label. In order to use the USDA organic seal, the final product has to follow strict production, handling, and labeling standards, such as soil quality, animal raising practices, and pest control. Also, due to public opinion in the 1990s, GMOs and biosolids are banned from being accepted as organic (Connor et al., 2004; McEvoy, 2016). That being said, many consumers still do not fully understand what the label entails, because food does not have to necessarily be healthy to be organic; food that contains high fructose corn syrup made with synthetic materials—which has shown to be extremely unhealthy (What is High Fructose, 2025)—can be labeled as organic because the synthetics are not incorporated into the final product (Hansen, 2013). This shows how consumers can be misled into believing organic products are healthier than they really are. In a press conference, the Secretary of Agriculture stated, 'Let me be clear about one thing: the organic food label is a marketing tool. It is not a statement about food safety or quality' (Cohrssen & Miller, 2016). Even though consumers might not fully understand the label, they have trust in the government to make sure that their food is safe; as an example, since the government began regulating organics, the organic food sector has nearly tripled in revenue in a decade (Carlson, 2016). While in the US, consumers might trust the organic label, that's not

always the case in other areas of the world. According to an EU study, only around half of participants recognized the EU organic label (Gorton et al., 2018).

Organic Food Claims

Organic food products are often marketed with claims that they promote better health and help prevent diseases. A 2017 European Commission classified these health claims into three types: functional health claims, risk reduction claims, and child development health claims. (European Commission, 2017, as cited in Küster & Vila, 2017, pp. 66-67). Understandably, there has been lots of research to prove these claims, for example in a 2025 study by the Mayo Clinic, they found that subjects who consumed more organic food had a decreased risk of developing a type of cancer called non-Hodgkin Lymphoma, while a massive French-Belgian study found that a higher consumption of organic food correlated with lower obesity risk (Kesse-Guyot et al., 2013). On a note related to the child development claim, a study by Torjusen et al. (2014), found that mothers who consumed more organic vegetables had a lower risk of developing pre-eclampsia. Additionally, organic meat and dairy products have been shown to contain higher levels of omega-3 fatty acids, which support heart and brain health (Średnicka-Tober et al., 2016). Looking at the results together, these studies show a consistent association between the consumption of organic food and overall health. However, along with the actual effect of the food, consumer perception of the food also has a big impact: McNeil (2025) found that Gen Z consumers especially associate organic labels with health and well-being. Ultimately, the strong association between organic foods and health not only influences individual consumption choices but also reflects wider concerns about sustainability and environmental impact, and regardless of whether these health benefits are fully understood by consumers, these associations likely contribute to the strong perception that organic foods are healthier than conventional alternative
Psychological Drivers Behind Organic Food Consumption

When considering the influence of the Organic label on consumer psychology, an important idea is the Halo Effect that organic products create among consumers. This can be seen throughout the world where an increasingly health-conscious populace is driving demand for organic versions of existing products (Wilkinson, 2025). This shows how the organic label can influence consumers to believe that an unhealthy product is healthy. This can be seen in a study from Mexico, where subjects were given one brownie with the organic label, and an identical brownie without the label. In almost all the trials they said the organic brownies were more nutritious and tasted healthier (Ruiz et al., 2024). Similarly, another study on organic wine found that consumers link organic to guilt-free pleasure, and increase their consumption (Petit et al., 2023, pp. 3046-3060). There are many psychological drivers behind purchases based on labels, such as moral satisfaction, self-image, and perceived health benefits. According to a 2022 study, FOPNLs and package design influence consumers by affecting their cognitive, biological, and cultural receptors, thereby generating an effect on consumer eating and purchasing behavior (Penzavecchia et al., 2022, pp. 3037-3051). There have been multiple research studies showing this, according to Wang et al. (2023), the main cause for purchasing and consuming organic food

was perceived health benefits and health awareness, however, another significant cause was maintaining social image. Another significant cause is the "warm glow" effect; which according to a study from Indonesia, had an overwhelming effect on purchasing behavior (Cahyasita et al., 2025, pp. 71-84). In conjunction with the "warm glow" effect, modern consumers, when buying meat, want to know how the animal was raised, how it was killed, and what happened to the meat after (Bredar, 2025). This all leads back to the main idea that buyers associate price with health, which according to Rao & Monroe (1989) is not irrational, as the forces of supply and demand would lead to a natural ordering of products on a price scale; leading to a strong association between price and quality (pp. 351-357). Another important psychological driver behind purchasing choices is marketing and presentation of the product. Specifically the design of the package, this can be seen in the previous study about wine, where the researchers found that consumers viewed organic red labeled drinks as more unhealthy compared to organic green labeled drinks (Petit et al., 2023, pp. 3046-3060). A common theme among studies of this nature seemed to be that consumers thought cooler colors were healthier; in a similar study, consumers perceived cool tones as healthier and sustainable (Richetin et al. 2022). Similarly, another common association is between darker colors and authenticity, in a study by Marozzo & Donato (2025), they found that dark brown olive oil bottles were perceived as healthier and more authentic compared to green or transparent ones (pp. 1-23). This shows how packaging design is an important factor in the psychological response to organic labels. Another psychological factor is the interaction between the label and branding strategies. This can be seen in a 2021 study where 39% of participants said that advertising and reliability played a role in their purchase decisions of organic foods (Chowdhury et al.). This ties into the importance of brand reputation; in a 2012 study, Larceneux et al. found that on products from brands with high reputation, the influence of the label was minimal, and vice versa (pp. 85-104). Overall the evidence shows that there are many psychological drivers behind purchase decisions influenced by the organic label. However, while the psychological drivers behind organic consumption are well documented, few studies compare whether these effects vary across demographic groups, leaving much uncertainty about who is most influenced by the label. Overall, these psychological associations between organic labels, health, and moral satisfaction may help explain why many consumers are willing to pay higher prices for organic products.

Consumer Justification of Organic Food Purchases

When discussing consumers' justifications for paying higher prices for organic products, research often focuses on their WTP. According to a 2015 study, the organic meat market is projected to grow by over a billion dollars over the next year. This growth is unprecedented because just a decade ago, organic meat was considered a luxury, now it's an everyday purchase (Bredar, 2015). This shows how consumers are much more willing to pay higher prices for organic products. In a 2023 study, the researchers found that on average, 88.5% of consumers in their study demonstrated a WTP for a price premium ranging from 5.6% up to 91% for healthier products (Alsubhi et al.). Similarly, in a 2017 study, the researchers found that 91.4% of the

participants they studied showed a WTP for a price premium ranging from 50% to 400% for products with the organic label (Dolgopolova & Teuber, pp. 333-352). Extending this pattern of consumer preference, Ghazanfari et al. found that a price premium does not significantly affect the consumer's willingness to buy organic products (2024, p. 6895). While all this evidence points to the fact that many consumers are willing to pay more for organic food, often there are problems with the surveying methods and thus the results. According to a 2020 study, the main problem with surveys on organic food is that people express higher positive emotions to organic food and negative emotions to conventional food in the survey. When, really, they often have equal emotions for both. Their results proved this: of 76% of participants who intended to buy organic food in the survey, only 13% actually bought organic food (Ismael & Ploeger, p. 79). Ultimately, the evidence suggests that although consumers express a strong WTP more for organic products, this willingness often reflects perceived health benefits rather than consistent purchasing habits. Despite extensive research on consumer WTP and attitudes toward organic foods, relatively few studies directly compare how demographic characteristics influence how consumers justify paying price premiums for organic products.

Assumptions

In order to conduct this study, it is assumed that the organic label influences consumer purchasing behavior by shaping perceptions of health, safety, and quality (Hansen, 2013; Cohrssen & Miller, 2016).

Justification

Consumer understanding of the USDA Organic label is often limited or distorted, with many confusing it with general health or “natural” claims (Hansen, 2013; Cohrssen & Miller, 2016; Diekman et al., 2023). Despite this, consumers continue to pay significant price premiums, motivated by perceived health, safety, or ethical benefits (Carlson, 2016; Dolgopolova & Teuber, 2018; Cahyasita et al., 2025). Research also shows that motivations and trust in the label vary across demographic groups such as age, income, and education (Kesse-Guyot et al., 2013; McNeil, 2025; Chowdhury et al., 2021). However, while many studies examine willingness to pay or general attitudes toward organic foods, relatively few directly compare how different demographic groups perceive price premiums or justify paying higher prices for organic products. Thus creating the gap this study aims to fill.

Research Question

While past research has focused mostly on consumer trust in organics and WTP for them, there has been limited to no evidence on how these behaviors differ amongst demographic groups. To address this gap in research, this study examines how the organic label influences the overall perception of products by consumers of various demographic groups.

Methodology

Purpose

The purpose of this study was to examine whether demographic factors influence WTP a price premium for organic food. Previous research (Keese-Guyot et al., 2013; Chowdhury et al., 2021) has suggested that characteristics such as age, gender, and income may influence organic food purchasing behavior. Therefore, this study analyzed whether demographic variables are associated with WTP for organic products and the perceived value of the organic label.

Data Collection Procedure

The survey consisted of two sections: (1) price estimation for organic and non-organic products and (2) perceived value of the organic label.

(1) To estimate the WTP, the participants were shown images of organic and non-organic versions of three products (ground beef, eggs, and mayonnaise). Using a method adapted from Richetin et al., participants estimated the price of each product using multiple-choice options or entering their own estimate. Responses were coded dichotomously (1 = price overestimation, 0 = no overestimation). Scores were summed to produce organic and non-organic overestimation totals (range 0–3). An Overestimation Difference (OD) score was calculated using the formula:

OD Score = (Total Organic Overestimation Score) − (Total Non-Organic Overestimation Score).

Positive values indicate greater overestimation of organic products; negative values indicate greater overestimation of non-organic products.

(2) Perceived value was measured using three Likert-scale questions.

Likert Scale Questions from the Data Collection Tool

Based on your best estimate, how often do you think about the organic label on foods when purchasing them?
Based on your best estimate, how often do you buy organic food?
Do you think the presence of the organic label of food makes it worth paying a price premium?

Participants gave a score of 1-5 on each question. During analysis, the average of the three scores were taken to create an overall “perceived value score” (PV Score) that was then used for descriptive and inferential statistics with the different demographic groups. To assess the reliability of the PV score, Cronbach's alpha was calculated using the three survey questions mentioned above. The analysis indicated good internal consistency ($\alpha = 0.82$), suggesting that the items measured a common underlying construct and were appropriate to combine into a composite score.

Method Justification

During planning, a qualitative approach was considered for the price perception section, where data would be collected through an interview and coded accordingly; however, the current method was chosen because it was already researched and reviewed by Richetin et al., and therefore I could use their questioning method and pricing scales for my data collection. For the value perception section, a qualitative approach was also considered; however, the current method was chosen because it could show correlation between variables using statistics rather than hinting at possible correlations. Lastly for the demographics section, the APA styling guide was used to create the demographic questions.

Participants

The target population was adults aged 18 or older. Participants were gathered using voluntary response sampling through flyers (as well as SurveyCircle and SurveySwap), postings on social media (Twitter, Facebook, LinkedIn, and Reddit), and reaching out to teachers in school. Finally, a $25 Amazon gift card raffle was held as an incentive with completion of the survey.

Survey Design

The survey (presented in **Appendix A**) began with a consent form (presented in **Appendix B**). Only participants who consented and 18+ were allowed to continue.

After these initial questions, participants were shown three side by side images of one organic and one non-organic version of the same product (mayo, ground beef, and eggs). Then, they were asked to estimate the price of each item.

Next, participants completed 3 likert-scale questions measuring their perceived value of the organic label.

Lastly, participants provided demographic information (age range, gender, ethnicity, income range, education level, location, etc.) to show how people of different demographic groups value the presence of the organic label differently

Finally, the participants were presented with a debriefing form (presented in **Appendix C**) which allowed them to permit the usage of their data in the results and for further analysis in this study. The survey was available and collected data from December 16th, 2025 until February 17th, 2026.

Ethical Considerations and Institutional Review Board

This research study, when reviewed by the IRB, was approved with no further modifications. To ensure the privacy and confidentiality of each participant, no information such as email addresses, phone numbers, or names were taken, so no participant could be linked to a specific set of answers. However, participants did voluntarily provide their email addresses if they wished to participate in the gift card raffle. Finally, all data was secured in the researchers google drive, with the project director being the only other person with access.

Findings

Demographics

A total of 313 responses were recorded. 8 were excluded for lack of consent or being under 18, leaving 305 participants. Some analyses used smaller samples due to missing responses. Although participants gave location and education information, many of the participants did not respond and the responses were extremely varied, with groups having less than 5 responses in the location category in particular. The survey examined price perception and perceived value of the organic label. Conditions for inference were assessed and assumed to be met (further information in **Appendix D)**. All tests used $\alpha = 0.05$.

Price Perception

Price perception was analyzed using the Overestimated Difference (OD) score described earlier in the methodology section.

Overall Label Bias

To assess the overall overestimation bias, the OD scores were analyzed.

Table 1

Descriptive Statistics for Price Perception Difference Scores

Sample Size	Sample Mean OD Score	Sample Standard Deviation
$n = 296$	$\bar{x} = -0.314$	$s = 0.931$

To test for an overall overestimation bias, a one sample t-test for the mean difference was conducted testing the sample against 0.

Table 2

One Sample t-Test for Mean Difference on Overall Label Bias

μ = the true mean difference in overestimation scores (organic score - non-organic score)

Hypotheses	t-Statistic	p-value	Effect Size (Cohen's *d*)
Ho: $\mu = 0$, Ha: $\mu \neq 0$	$t(295) = -5.802$	p-value ≈ 0	$d = 0.34$

Since $0 < 0.05$, we reject Ho, and have convincing evidence that people generally overestimate the price of non-organic products. Now that it had been established that there was an overall labeling bias, the next step was to see if the trend applied between different demographic groups (gender, age, ethnicity, etc).

Overall Label Bias by Gender

To assess overestimation bias by gender, OD scores were analyzed (n = 286).

Table 3

Descriptive Statistics for OD Scores by Gender

Gender	Sample Size	Sample Mean OD Score	Sample Standard Deviation
Female	n = 203	$\bar{x}$ = -0.285	s1 = 0.910
Male	n = 83	$\bar{x}$ = -0.422	s2 = 0.952

Note: 19 participants were excluded because they either did not answer the questions or answered prefer not to say

To test whether the mean OD scores differed significantly between genders, a two-sample t-test for $\mu 1 - \mu 2$ was conducted.

Table 4

Two sample t-Test for Difference in Mean OD Scores by Gender

$\mu 1 - \mu 2$ = the true difference in mean OD scores (Female - Male)

Hypotheses	t-Statistic	p-value	Effect Size (Cohen's *d*)
Ho: $\mu 1 = \mu 2$, Ha: $\mu 1 \neq \mu 2$	t(146.417) = -3.495	p-value = 0.000628	d = 0.15

Since 0.000628 < 0.05, we reject Ho, and have convincing evidence that there is labeling bias between genders, with males overestimating the price of non-organic products more often than females.

Overall Label Bias by Age

To assess overestimation bias by age, OD scores were analyzed.

Table 5

Descriptive Statistics for OD Scores by Age Group

Age Group	Sample Size	Sample Mean OD Score	Sample Standard Deviation
18-24	n = 110	$\bar{x}$ = -0.236	s = 0.918
25-34	n = 73	$\bar{x}$ = -0.370	s = 0.979
35-44	n = 49	$\bar{x}$ = -0.510	s = 1.003
45-79	n = 62	$\bar{x}$ = -0.242	s = 0.824

	N = 294	Note: 11 participants were excluded because the did not answer the questions or answered “prefer not to say”

To determine if the mean OD scores differed significantly across age groups, a one-way ANOVA was conducted.

Table 6

One-Way ANOVA for OD Scores by Age Group

μ1, μ2, μ3, μ4 = the true mean OD scores for each age group

Hypotheses	F-Statistic	p-value	Effect Size (η^2)
Ho: μ1 = μ2 = μ3 = μ4, Ha: at least one is different	F(3, 290) = 1.1	p-value = 0.35	$\eta^2 = 0.012$

Since 0.35 > 0.05, we fail to reject Ho, and do not have convincing evidence that there is a significant difference in overall labeling bias across age groups.

Overall Label Bias by Ethnicity

To assess overestimation bias by ethnicity, OD scores were analyzed.

Table 7

Descriptive Statistics for OD Scores by Ethnicity

Ethnicity	Sample Size	Sample Mean OD Score	Sample Standard Deviation
White	n = 164	$\bar{x}$ = -0.25	s = 0.897
Black	n = 23	$\bar{x}$ = -0.522	s = 0.994
Asian	n = 90	$\bar{x}$ = -0.356	s = 0.891
Hispanic	n = 5	$\bar{x}$ = -0.400	s = 0.894
Jewish	n = 2	$\bar{x}$ = -1.000	s = 1.414
Pacific Islander	n = 2	$\bar{x}$ = -0.500	s = 0.707
	N = 277	Note: the N = 277 excludes the 3 groups whose sample size isn’t large enough for inferential stats, and adds an extra 5 for people who identified multiple ethnicities	

To determine if the mean OD scores differed significantly across ethnic groups, a one-way ANOVA was conducted.

Table 8

One-Way ANOVA for OD Scores by Ethnic Group

μ1, μ2, μ3 = the true mean OD scores for each age group			
Hyptotheses	**F-Statistic**	**p-value**	**Effect Size (η^2)**
Ho: μ1 = μ2 = μ3, Ha: at least one is different	F(2, 274) = 1.11	p-value = 0.33	$\eta^2 = 0.009$

Since 0.33 > 0.05, we fail to reject Ho, and we do not have convincing evidence that there is a significant difference in labeling bias across ethnic groups.

Overall Label Bias by Income

To assess overestimation bias by income, OD scores were analyzed.

Table 9

Descriptive Statistics for OD Scores by Income

Income	**Sample Size**	**Sample Mean OD Score**	**Sample Standard Deviation**
$0 - $14,999	n = 46	$\bar{x} = -0.217$	s = 0.786
$15,000 - $29,999	n = 35	$\bar{x} = -0.029$	s = 0.747
$30,000 - $59,999	n = 60	$\bar{x} = -0.200$	s = 1.054
$60,000 - $99,999	n = 41	$\bar{x} = -0.585$	s = 1.140
$100,000 - $199,999	n = 39	$\bar{x} = -0.359$	s = 0.959
$200,000+	n = 29	$\bar{x} = -0.276$	s = 0.649
	N = 250	Note: 55 participants were excluded because they either did not answer or answered with “prefer not to say”	

To determine whether the mean OD scores differed significantly across income groups, a one-way ANOVA was conducted.

Table 10

One-Way ANOVA for OD Scores by Income Group

μ1, μ2, μ3, μ4, μ5, μ6 = the true mean OD scores for each age group			
Hyptotheses	**F-Statistic**	**p-value**	**Effect Size (η^2)**
Ho: μ1 = μ2 = μ3 = μ4 = μ5 = μ6, Ha: at	F(5, 244) = 1.11	p-value = 0.16	$\eta^2 = 0.019$

least one is different

Since 0.16 > 0.05, we fail to reject Ho, and do not have convincing evidence that there is a significant difference in labeling bias across income groups.

Perceived Value of Organic Label

Perceived value of the organic label was analyzed using the Perceived Value (PV) score described in the methodology.

Perceived Value of Organic Label by Gender

To assess value perception by gender, PV scores were analyzed. (n = 292).

Table 11

Descriptive Statistics for PV Scores by Gender

Gender	Sample Size	Sample Mean PV Score	Sample Standard Deviation
Female	n = 207	$\bar{x}$ = 3.161	s1 = 0.846
Male	n = 85	$\bar{x}$ = 3.286	s2 = 0.925

Note: 13 participants were excluded because they either did not answer or answered with "prefer not to say"

To test whether the mean PV scores differed significantly between genders, a two-sample t-test for μ1 - μ2 was conducted.

Table 12

Two sample t-Test for Difference in Mean PV Scores by Gender

μ1 - μ2 = the true difference in mean P scores (Female - Male)

Hypotheses	t-Statistic	p-value	Effect Size (Cohen's *d*)
Ho: μ1 = μ2, Ha: μ1 ≠ μ2	t(144.656) = 1.075	p-value = 0.284	*d* = 0.14

Since 0.284 > 0.05, we fail to reject Ho, and do not have convincing evidence that there is an overall difference in the value perception of the organic label between genders.

Perceived Value of Organic Label by Age
To assess value perception by age, PV scores were analyzed.

Table 13

Descriptive Statistics for PV Scores by Age Group

Age Group	Sample Size	Sample Mean PV Score	Sample Standard Deviation
18-24	n = 111	$\bar{x}$ = 3.089	s = 0.847
25-34	n = 74	$\bar{x}$ = 3.324	s = 0.788
35-44	n = 50	$\bar{x}$ = 3.533	s = 0.979
45-79	n = 63	$\bar{x}$ = 3.085	s = 0.899
	N = 298	Note: 7 participants were excluded because they either did not answer or answered with "prefer not to say"	

To determine if the mean PV scores differed across age groups, a one-way ANOVA was conducted.

Table 14

One-Way ANOVA for PV Scores by Age Group

μ1, μ2, μ3, μ4 = the true mean PV scores for each age group

Hyptotheses	F-Statistic	p-value	Effect Size (η^2)
Ho: μ1 = μ2 = μ3 = μ4, Ha: at least one is different	F(3, 294) = 3.43	p-value = 0.018	η^2 = 0.018

Since 0.018 < 0.05, we reject Ho, and have convincing evidence that the overall value perception of the organic label differs by age, with individuals aged 35-44 showing the highest value perception of the organic label. A post-hoc Tukey test was conducted to see which differences were the most significant, however, while there was an overall significant difference, the test revealed no statistically significant pairwise differences between age groups.

Perceived Value of Organic Label by Income

To assess value perception by income, PV scores were analyzed.

Table 15

Descriptive Statistics for PV Scores by Income

Income	Sample Size	Sample Mean PV Score	Sample Standard Deviation
$0 - $14,999	n = 46	$\bar{x}$ = 2.978	s1 = 0.850
$15,000 - $29,999	n = 36	$\bar{x}$ = 3.241	s2 =0.767
$30,000 - $59,999	n = 60	$\bar{x}$ = 3.233	s3 = 0.840

$60,000 - $99,999	n = 42	$\bar{x}$ = 3.230	s4 = 0.964
$100,000 - $199,999	n = 43	$\bar{x}$ = 3.411	s5 = 0.937
$200,000+	n = 29	$\bar{x}$ = 3.080	s6 = 1.003
	N = 256	Note: 49 participants were excluded because they either did not answer or answered with “prefer not to say”	

To determine if the mean PV scores differed across income groups, a one-way ANOVA was conducted.

Table 16

One-Way ANOVA for PV Scores by Income Group

μ1, μ2, μ3, μ4, μ5, μ6 = the true mean PV scores for each age group			
Hyptotheses	**F-Statistic**	**p-value**	**Effect Size (η^2)**
Ho: μ1 = μ2 = μ3 = μ4 = μ5 = μ6, Ha: at least one is different	F(5, 250) = 1.299	df2 = 250	η^2 = 0.025

Since 0.266 > 0.05, we fail to reject Ho, and do not have convincing evidence that there is a significant difference in the value perception of the organic label across income groups.

Perceived Value of Organic Label by Ethnicity
To assess value perception by ethnicity, PV scores were analyzed.

Table 17

Descriptive Statistics for PV Scores by Ethnicity

Ethnicity	Sample Size	Sample Mean PV Score	Sample Standard Deviation
White	n = 166	$\bar{x}$ = 3.040	s3 = 0.909
Black	n = 25	$\bar{x}$ = 3.560	s2 = 1.075
Asian	n = 96	$\bar{x}$ = 3.299	s1 = 0.721
Hispanic	n = 5	$\bar{x}$ = 3.667	s = 0.782
Jewish	n = 2	$\bar{x}$ = 2.167	s = 0.236
Pacific Islander	n = 2	$\bar{x}$ = 3.167	s = 0.236
	N = 287	Note: the N = 287 excludes the 3 groups whose sample size isn’t	

	large enough for inferential stats, and adds and extra 7 for people who identified multiple ethnicities

To determine whether the mean OD scores differed significantly across ethnic groups, a one-way ANOVA was conducted.

Table 18

One-Way ANOVA for PV Scores by Ethnic Group

μ1, μ2, μ3 = the true mean PV scores for each age group

Hyptotheses	F-Statistic	p-value	Effect Size (η^2)
Ho: μ1 = μ2 = μ3, Ha: at least one is different	F(2, 283) = 5.26	p-value = 0.006	$\eta^2 = 0.036$

Since 0.006 < 0.05, we reject Ho, and we have convincing evidence that there is a significant difference in the value perception of the organic label across ethnic groups, with overall Black participants showing the highest perceived value. A post-hoc Tukey test was conducted to see which differences were the most significant.

Table 19

Tukey Test Results Based on One-Way ANOVA for PV Scores by Ethnic Group

Group Comparisons	q-Statistic	p-value
Asian - Black	q = 1.96	p-value = 0.30
Asian - White	q = 2.08	p-value = 0.049
Black - White	q = 4.03	p-value = 0.018

Since the Tukey post-hoc test showed that both the Black-White comparison (p = 0.018 < 0.05) and the Asian-White comparison (p = 0.049 < 0.05) were statistically significant at α = 0.05, we have convincing evidence that Black and Asian participants reported significantly higher perceived values of the organic label compared to White participants. The Asian-Black comparison (p = 0.30 > 0.05), however, was not significant.

Correlations between Label Bias and Perceived Label Value

The final analysis examined the relationship between OD and PV scores.

Table 20

Descriptive Statistics for PV and OD Scores

Score	Sample Size	Sample Score Mean	Sample Standard Deviation
OD Score	n = 295	$\bar{x}$ = -0.308	s = 0.928
PV Scores	n = 295	$\bar{x}$ = 3.202	s = 0.873

Note: 10 participants were excluded because they either did not have an OD or PV score or they had only one of them

A Pearson correlation tested for a linear relationship between OD and PV scores.

Table 21

Pearson Correlation Between OD and PV Scores

ρ = The true population correlation coefficient				
Hypotheses	**r**	**df**	**p-value**	**Effect Size (r^2)**
Ho: ρ = 0, Ha: ρ ≠ 0	r = -0.0682	df = 293	p-value = 0.24	r^2 = 0.00465

Since 0.24 > 0.05, we fail to reject Ho, and do not have convincing evidence that there is a significant linear relationship between the Overestimation Difference score and Perceived Value score. However, based on r = -0.0682, we can say there is a very weak negative relationship between the two variables.

Analysis

This analysis connects the study's findings on price perception and perceived value of organic food to the existing research on the organic label and its effects, with a focus on psychological biases and demographic variations. The results overall support, extend, and challenge prior literature, by showing the distinction between perception and justification.

As mentioned before, overall participants significantly overestimated the price of non-organic food showing the existence of the halo effect (Andreoni, 1990, pp. 464-477). As discussed in the previous literature, many consumers associate organic products with higher quality and health, despite its actual use as a marketing tool (Hansen, 2013; Cohrssen & Miller, 2016). This misinterpretation of the label appears to extend to price perception as well, suggesting that consumers do not evaluate products objectively, but instead rely on assumptions about the organic label (Ruiz et al., 2024; Richetin et al., 2022). However, as previous research has shown, this bias often operates in a specific direction, with participants not simply valuing organic products more, but overestimating the cost of non organic alternatives. This distinction, surprisingly, is not very prevalent in prior research, which tends to focus more on the WTP rather than the misperception of prices (Dolgopolova & Teuber, 2018; Alsubhi et al., 2023). Therefore the findings of this study show that consumer decisions may be driven less by the valuation of the organic products themselves, but more by a distorted price perception of the non-organic

alternatives. This aligns with research showing that consumers use price as a proxy for quality and rely on simplified marketing cues when comparisons, rather than evaluating the actual value of the product independently (Rao & Monroe, 1989; Richetin et al., 2022). Overall this helps explain why consumers may justify paying more for organic products even when the value is lesser compared to non-organic alternatives. While most demographic factors did not significantly affect this bias, gender differences were observed, with males showing a greater tendency to overestimate the price of non-organic products. A possible explanation for this overestimating could be that while not significant, male respondents did value the organic label more than female respondents, which could provide a reasoning behind why they overestimated the price of non-organic food, as they would have a higher value perception of organic products, causing them to overestimate the price of non-organic alternatives. These results partially align with previous research showing that consumer perception is influenced by psychological and social factors (Wang et al., 2013). However, the lack of clear variation across age, income, and ethnic groups contrasts with previous studies that emphasize the demographic variation in organic attitudes. Overall this suggests that price misperception may be a more universally subconscious cognitive bias, rather than one shaped by demographic identity.

In contrast to price perception, the perceived value of the organic label showed much greater variability. While previous research highlights strong WTP for organic products caused by the perceived health benefits and ethical considerations (Dolgopolova & Teuber, 2018; Cahyasita et al., 2025), this study found only moderate perceived value scores (with the averages being around a 3), This supports the findings from Ismael & Ploeger, who found that consumers often express positive attitudes towards organic products in theory, but do not consistently follow through on those attitudes into strong value perception of purchasing behavior (2022). Unlike price perception, perceived value showed some more demographic variation. Age had a statistically significant effect, with middle-aged participants reporting higher perceived value, which challenged previous research, which showed that on average younger generations often associate a higher value to the organic label (McNeil, 2025). Overall, though, further analysis shows no clear pairwise differences. On the other hand, ethnicity showed significant differences with Black and Asian participants reporting higher perceived value than White participants. This aligns with previous literature suggesting the cultural and social factors influence how consumers interpret and assign meaning to the organic label, particularly in relation to health and ethical concerns (Kesse-Guyot et al., 2013; Chowdhury et al., 2021). These findings overall suggest that while price misrepresentation is relatively consistent across groups, value justification is slightly more dependent on social and cultural factors. This distinction supports previous research, which finds demographic differences in WTP; even when the overall perceptions of organic products were similar (Wang et al., 2023; Cahyasita et al., 2025)

On an interesting note, there was no significant relationship between price misrepresentation and perceived value. This challenges the assumption, supported by prior research, that higher perceived value directly drives WTP (Rao & Monroe, 1989; Alsubhi et al., 2023). Instead, the results suggest that price perception and justification may operate as separate

psychological processes–one driven by a general cognitive bias across demographic groups, and the other by a more conscious understanding and interpretation of the value influenced by the social and demographic factors (Penzavecchia et al., 2022).

Overall, the results of this study supports, extend, and challenge the existing literature by demonstrating that the effects of the organic label influence consumers in multiple ways. While prior research emphasizes willingness to pay and perceived health benefits, this study reveals that label-driven misperception of non-organic products plays an equally important role (Hansen, 2013; Dolgopolova & Teuber, 2018). Additionally, the findings highlight that demographic differences are more relevant in value justification than in initial perception, offering a more nuanced understanding of how consumers interpret organic labels.

Limitations

One possible limitation of this study is the use of voluntary response sampling, which can introduce response bias, as individuals with stronger opinions about organic food are more likely to participate. This could skew the results by overrepresenting individuals who are more engaged or opinionated about organic products. Additionally, that lack of random sampling to acquire participants means the results of this study can not be generalized to the broader population of adults.

Another limitation is that the study relies on self-reported estimates rather than actual purchasing behavior. Prior research suggests that consumers often express stronger intentions to purchase organic products than they demonstrate in real-life situations. As a result, participants' reported perception of value may not fully reflect their actual purchasing decisions, limiting the validity of the findings. The measurement of the variables also presents a limitation. The Overestimation Difference (OD) score and Perceived Value (PV) score simplify complex consumers perceptions into simple numerical values, which could fail to fully capture the many reasoning behind the participants' choices, such as health beliefs, ethical motivations, or prior experience with the organic label. Additionally, some demographic subgroups–particularly within ethnicity– had relatively small sample sizes, and were not included in inferential analyses, which could reduce the reliability of comparisons and increase variability in the results.

Another possible limitation of this study can be seen with the OD Score. Due to how it's calculated, it is a bounded discrete variable with a range from -3 to 3. Additionally, it only shows the direction in which a respondent estimated the price of organic/non-organic food, not the magnitude to which they estimated the prices. As a result of this system, participants who slightly over/under estimated the price of a product receive the same score as participants who significantly over/under estimated the price of a product. Despite the OD score not being a continuous variable, parametric statistical tests like the one-way ANOVA and t-test were still used due to the large sample size and the general robustness of these statistical methods.

Finally, because this study is observation, rather than an experiment, it cannot establish causality. Therefore, while relationships between demographic factors and perceptions can be identified, it cannot be concluded that these factors directly cause differences in price perception

or perceived value. Despite these limitations, the study can still provide a meaningful insight into how demographic factors correlate with the perceived value of the organic label.

While the presence of voluntary response and non-random sampling limit the generalizability of the study, the sample will still include a diverse range of participants across many different demographic groups, allowing us to identify meaningful patterns and relationships within the data. Additionally, although causation cannot be established due to the observational design, identifying statistical significance between a demographic variable and the dependent variable can still help guide future experimental research and inform hypotheses for more controlled studies.

Conclusion

This study examined the extent to which the organic label influences consumers of different demographic groups to justify paying higher prices and their overall value perceptions of organic products. The findings suggest that the organic label may influence consumer perception on a broad societal level rather than on an individual level, as participants in the study as a whole significantly tended to overestimate the price of non-organic products regardless of their background. As to the value perception, the overall effect was more moderate; the findings suggest that it depends more on the societal and cultural factors that influence the individuals, and can vary by demographic group. Additionally, the lack of a significant relationship between price perception and perceived value suggest that these processes may operate independently, where the former could be more of a societal effect and the latter more of a subconscious individual effect. Overall, these findings have many important implications for consumer behavior and food labeling policy; organic labels may influence consumers' purchasing decisions, not by increasing the actual health of the product, but by distorting the value of conventional products. This raises concerns about consumer misunderstanding and highlights the importance of clearer labeling and consumer education. Additionally, marketers can leverage these perceptual biases, further increasing the disparity between actual value and marketed value.

Future Directions

Future research could create more accurate conclusions by using an experimental design rather than a survey in order to establish causal relationships between the demographic variables and price perception and value perception. Additionally, better understanding could be established by gathering real data on purchasing behavior rather than self-reported data, which would better show how the organic label influences real decision making. Expanding sample diversity and including additional variables, such as consumer knowledge or purchasing frequency, could also help further explain the role of demographic differences in shaping perception and value. Finally, future research could also examine the magnitude of price estimation in addition to the directional bias to create a more thorough understanding of the psychological process behind price estimation.

Works Cited

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow. Royal Economic Society, 100(401), 464–477. https://www.jstor.org/stable/2234133.

Alsubhi, M., Blake, M., Nguyen, T., Majmudar, I., Moodie, M., & Ananthapavan, J. (2023). Consumer willingness to pay for healthier food products: A systematic review. Obesity Reviews, 24(1) https://doi.org/10.1111/obr.13525.

Bredar, S. (2025). Natural and organic products in particular are in a position to capitalize on the farm-to-fork movement. The National Provisioner, 237(5). https://link.gale.com/apps/doc/A840726825/STOM?u=alli1510&sid=bookmark-STOM&xid=d468e3de.

Cahyasita, D., Irham, I., & Jamhari. (2025). Introducing "warm glow" as a key psychological motive on consumer willingness to consume organic food: A study of ethical consumption behaviour. The International Journal of Sociology of Agriculture and Food, 31(1), 71–84.https://doi.org/10.48416/ijsaf.v31i1.547.

Carlson, A. (2016). Investigating retail price premium for organic food. Economic Research Service, U.S.Department of Agriculture. https://www.ers.usda.gov/amber-waves/2016/may/investigating-retail-price-premiums-for-organic-foods.

Chowdhury, S., Meeru, A., Rahman, A. A. A., Anwarul Islam, K. M., Zayed, N. M., & Rajibul Hasan,K. B. M. (2021). An empirical study on the factors affecting organic food purchasing behavior in Bangladesh: Analyzing a few factors. Academy of Strategic Management Journal, 20(4). https://www.abacademies.org/articles/An-empirical-study-on-the-factors-affecting-organic-food-purchasing-behavior-in-bangladesh-analyzing-a-few-factors-1939-6104-20-4-815.pdf.

Cohrssen, J. J., & Miller, H. I. (2016). The USDA's meaningless organic label. CATO Institute.https://www.cato.org/regulation/spring-2016/usdas-meaningless-organic-label.

Diekman, C., McCargar, J., & Gibson, D. (2023). Misinformation and disinformation in food science and nutrition: Impact on practice. The Journal of Nutrition, 153(1), 3–9. https://doi.org/10.1016/j.jand.2022.13.1020.

Dolgopolova, I., & Teuber, R. (2018). Consumers' willingness to pay for health benefits in food products: A meta-analysis. Applied Economic Perspectives and Policy, 40, 333–352. https://doi.org/10.1093/aepp/ppx036.

Ghazanfari, S., Firoozzare, A., Covino, D., Boccia, F., & Palmieri, N. (2024). Exploring factors influencing consumers' willingness to pay healthy-labeled foods at a premium price. Sustainability, 16(16), 6895. https://doi.org/10.3390/su16166895.

Gorton, M., Tocco, B., Yeh, C.-H., & Hartmann, M. (2018). What determines consumers' trust in the EU's organic label? A cross-country comparison of the role of institutional trust and consumer knowledge. https://doi.org/10.22004/AG.ECON.273857.

Hansen, J. (2013). Interpreting food labels: Natural vs organic. American Society for Nutrition. https://nutrition.org/interpreting-food-labels-natural-versus-organic/.

Ismael, D., & Ploeger, A. (2020). Consumers' emotional attitudes towards organic and

conventional food: A comparative study of emotional profiling and self-reported methods. Foods, 9(1), 79. https://doi.org/10.3390/foods9010079.

Kanter, R., Vanderlee, L., & Vandevijvere, S. (2018). Front-of-package nutrition labelling policy: Global progress and future directions. Public Health Nutrition, 21(8), 1399–1408. https://doi.org/10.1017/S1368980018000010.

Kesse-Guyot, E., Péneau, S., Méjean, C., Szabo de Edelenyi, F., Galan, P., Hercberg, S., & Denis, L. (2013). Profiles of organic food consumers in a large sample of French adults: Results from the Nutrinet-Santé cohort study. PLOS ONE, 8(10), e76998. https://doi.org/10.1371/journal.pone.0076998.

Küster, I. (2017). Health/nutrition food claims and low-fat food purchase: Projected personality influence in young consumers. Journal of Functional Foods, 38(A), 66–67. https://doi.org/10.1016/j.jff.2017.08.046.

Larceneux, F., Benoit-Moreau, F., & Renaudin, V. (2012). Why might organic labels fail to influence consumer choices? Marginal labelling and brand equity effects. Journal of Consumer Policy, 35, 85–104. https://doi.org/10.1007/s10603-011-9186-1.

Marozzo, V., & Donato, C. (2025). The impact of packaging color and sustainable labels on perceived food healthfulness. Journal of Marketing Theory and Practice, 1–23. https://doi.org/10.1080/10696679.2025.2566310.

Mayo Clinic Staff. (n.d.). Organic foods: Are they safer? More nutritious? Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/organic-food/art-20043880.

McEvoy, M. (2016). Understanding the USDA organic label. United States Department of Agriculture.https://www.usda.gov/about-usda/news/blog/understanding-usda-organic-label.

McNeil, M. (2025). Millennials, Gen Z putting high value on organic's health benefits. Organic Trade Association. https://ota.com/about-ota/press-releases/younger-health-conscious-consumers-are-embracing-organic-ota-survey-shows.

Organic Trade Association. (2024). 2024 organic industry survey, Organic Trade Association, https://ota.com/resources/market-analysis/organic-industry-survey/2024-organic-industry-survey.

The Packer. (2021). 2020 sees record organic sales, OTA reports. https://www.thepacker.com/news/organic/2020-sees-record-organic-sales-ota-reports

Penzavecchia, C., Todisco, P., Muzzioli, L., Poli, A., Marangoni, F., Poggiogalle, E., Giusti, A. M., Lenzi, A., Pinto, A., & Donini, L. M. (2022). The influence of front-of-pack nutritional labels on eating and purchasing behaviors: A narrative review of the literature. Eating and Weight Disorders, 27(8), 3037–3051. https://doi.org/10.1007/s40519-022-01507-2.

Pfaff, S., (2025). Growth in U.S. organic marketplace accelerated in 2024. Organic Trade Association,

https://ota.com/about-ota/press-releases/growth-us-organic-marketplace-accelerated-2024
Petit, O., Wang, Q. J., & Spence, C. (2023). The “healthy = (un)tasty” intuition concerning colour in organic wine labels. Journal of Consumer Behaviour, 23(6), 3046–3060. https://doi.org/10.1002/cb.2394.
Rao, A. R., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers’ perceptions of product quality. Journal of Marketing Research, 26, 351–357. https://doi.org/10.1177/002224378902600309.
Richetin, J., Caputo, V., Demartini, E., Conner, M., & Perugini, M. (2022). Organic food labels bias food healthiness perceptions: Estimating healthiness equivalence using a discrete choice experiment. Appetite, 172, 105970. https://doi.org/10.1016/j.appet.2022.105970.
Rude, J., (2022). Top health strategies by generations, Euromonitor, https://www.euromonitor.com/article/top-health-strategies-by-generations.
Ruiz, F., Muñeton, G., Cárdenas, M. V., & Restrepo-Morales, J. A. (2024). Organic food labeling and consumer perception: Empirical evidence. https://doi.org/10.30878/ces.v32n0a4.
Średnicka-Tober, D., Barański, M., Seal, C., Sanderson, R., Benbrook, C., Steinshamn, H., … Leifert, C. (2016). Composition differences between organic and conventional meat: A systematic literature review and meta-analysis. British Journal of Nutrition, 115(6), 994–1011. https://doi.org/10.1017/S0007114515005073.
Torjusen, H., Brantsaeter, A. L., Haugen, M., Alexander, J., Bakketeig, L. S., Lieblein, G., Stigum, H., Naes, T., Swartz, J., Holmboe-Ottesen, G., Roos, G., & Meltzer, H. M. (2014). Reduced risk of pre-eclampsia with organic vegetable consumption: Results from the prospective Norwegian Mother and Child Cohort Study. BMJ Open, 4(9), e006143. https://doi.org/10.1136/bmjopen-2014-006143.
Wang, J., Xue, Y., & Liu, T. (2023). Consumer motivation for organic food consumption: Health consciousness or herd mentality. Frontiers in Public Health, 10, 1042535. https://doi.org/10.3389/fpubh.2022.1042535.
Wilkinson, F. (2025). Private-label offerings meeting consumer demands for value: More than half of Gen Z shoppers say they “always/frequently” choose where they shop based on store brands. The National Provisioner, 237(6). https://link.gale.com/apps/doc/A854113349/STOM?u=alli1510&sid=bookmark-STOM&xid=1c0ae3e2.

The Influence of Cognitive Versus Experiential Knowledge on Investing Behaviors Across Demographics By Mahika Shah

Abstract

This study examines how cognitive versus experiential knowledge of financial crises shapes investment behavior across generations, with a focus on Generation Z's high participation in high-risk investments such as cryptocurrency and meme stocks. With speculative investing among young Americans steadily increasing, behavioral patterns associated with social media influence, financial nihilism, and herd mentality appear with views on risk and investment decisions. While previous research has identified generational differences in investment behavior, few studies have directly compared the behavioral effects of cognitive and experiential financial knowledge across matched demographic groups. To address this gap, a population-weighted online survey of 163 American adults was conducted in February 2026, analyzing investment attitudes and behaviors across six age groups using chi-squared tests of independence. The results showed that age, which reflects experiential knowledge, was the strongest predictor, though the mechanism behind this association remains unclear from the data. It produced significant results in 14 of 17 questions, while gender and current household income did not show meaningful effects. This suggests that age is the strongest demographic predictor of investment behavior, though whether this reflects experiential knowledge, wealth accumulation, life stage, or other age-correlated factors cannot be determined from these data. Younger investors may be more vulnerable, as limited real-world experience is associated with greater risk-taking in this dataset. This points to the need for learning programs that give experiential knowledge, not just traditional financial lessons. Future studies should examine how different demographics or regions influence investing behaviors.

Keywords Generation Z, experiential knowledge, cognitive knowledge, investment behavior, financial crises, behavioral finance, risk tolerance

1. Literature Review

1.1 Cognitive Knowledge and Investment Behavior

Cognitive knowledge refers to information about investing gained through reading or studying, instead of direct market participation.

1.1.1 The Description-Experience Gap

Behavioral finance is separated in two parts: cognitive and experiential knowledge. This difference is seen in the "description-experience gap," which suggests that decision-making from cognitive knowledge tends to make individuals overestimate the likeliness of rare events such as crashes, whereas decision-making from experience causes individuals to underestimate those risks unless they have experienced them recently (Olschewski et al., 2023). In other words, an

individual who reads about risk and what they feel from experiencing it can lead to very different decisions.

1.1.2 Financial Literacy and Its Limits

Higher financial literacy can improve a person's ability to analyze potential investments and reduce the likelihood for falling victim to herd mentality. However, its influence tends to be outweighed by expectations of returns (Asri et al., 2024). Even individuals with enough financial literacy tend to knowingly participate in risky investing behaviors if the possible outcomes seem attractive enough (Birje et al., 2025). Importantly, learning more about finance can actually make individuals overconfident, causing them to underestimate risk and trade too much (Dhamma & Setyawati, 2026; Darwish, 2025). Comparative generational analysis further complicates this: Generation Z demonstrated less theoretical financial knowledge when compared to Generation X and Baby Boomers, but exhibited higher overall rates of market participation. This suggests that they follow a "learn by doing" approach acquired through finfluencers instead of traditional knowledge (Mucskova et al., 2025).

1.1.3 Herd Mentality

Social Learning Theory suggests that individuals follow a certain behavior based on their role model (Spohn, 2024). In digital investment contexts, this is seen as amplified herding, where Gen Z trading is more often driven by online communities than individual research (Anggiani et al., 2025). The GameStop rally of January 2021 is an example—a coordinated trade driven by social trends rather than the company's actual value. This makes Gen Z more prone to herd mentality due to cognitive knowledge which provides no emotional scarring.

1.1.4 Fear of Missing Out (FOMO)

Fear of Missing Out (FOMO) is another significant driver of risky behavior, with 50% of U.S. Gen Z investors reporting at least one FOMO-driven investment (66% in China; 55% in the U.K.) (CFA Institute, 2023). Importantly, Gen Z's FOMO functions differently than regular FOMO because it is driven by a fear of missing out on possible financial gains. Many Gen Z investors make high risk investments with the belief that traditional investments cannot provide the same return (Spohn, 2024).

1.1.5 Gamification and Investing

Mobile trading apps have made it a lot easier for people to start investing, but also make investing seem more like a game instead of something serious, creating a negative view of the economy and implementing a unique environment for risky investing. Features such as confetti animations, constant notifications, and simple designs encourage impulsive decisions instead of careful thinking (The Echo, 2026). Research displays a strong correlation between frequent gambling behavior and speculative investing among Gen Z with 70% of frequent gamblers holding cryptocurrency positions and 38% holding NFTs (CFA Institute, 2023).

1.1.6 Digital Financial Socialization and Finfluencers

Financial socialization, how people develop the skills and attitudes to participate in markets, was traditionally passed down through parents and institutions (Sundarasen et al., 2026). For Generation Z, this way of learning has been largely replaced by digital platforms. While parents remain one of the most trusted sources for financial advice (cited by 27% of Gen Z investors), they are no longer the top source where Gen Z actually learns about investing (CFA Institute, 2023). Instead, 48% of Gen Z investors report learning from social media and 47% from internet searches, while only 13% cite formal knowledge as their most helpful source (HRB Tax Group, 2024).

This gap has now been filled by "influencers" who are social media content creators who give financial knowledge without any professional credentials. Their influence works through parasocial relationships (one-sided emotional bonds where followers feel a personal connection that makes the creator's advice seem more worthy than that of a traditional expert) (Hii & Ong, 2025). Research using the Elaboration Likelihood Model shows the Gen Z analyzes influencer content on both its content and perceived likability (Hii & Ong, 2025). Since platforms prioritize engagement over accuracy, algorithms repeatedly show misleading success stories that increase FOMO among young viewers (The Echo, 2025; Ahmmad et al., 2025). Importantly, all of this acts as a form of information consumption rather than lived market experience.

1.2 Experiential Knowledge and Investment Behavior

Experiential knowledge is gained through lived experiences of financial events instead of just reading about them, which creates a different effect on how an investor views certain investment choices compared to cognitive knowledge.

1.2.1 The Experience Effect and Affective Scarring

Malmendier's "experience effect" demonstrates that personal financial experiences impact an individual's investing beliefs, risk attitudes, and market participation for decades (Malmendier, 2021). Generations who lived through financial crises such as the 2000 Dot-com collapse or the 2008 Global Financial Crisis carry a form of "affective scarring", where individuals are still affected by the emotional weight of financial crises long after markets have recovered (Malmendier, 2021). This emotional scarring is a type of knowledge that cannot be easily replicated by reading about a crisis.

The Experience-Based Learning model (EBL) explains why these generational differences continue to exist. Unlike a Full Bayesian Learner (FBL) who accounts for all historical data and weighs them equally, an EBL places more importance on personal experience, especially recent ones (Malmendier et al., 2017). For younger investors with little market history, any single event such as a meme stock rally, makes up a much larger portion of their total experience, having a bigger impact on recent trends and causing them to be more optimistic during booms (Price, 2024).

1.2.2 Historical Amnesia

Gen Z's reduced hesitancy towards financial risk is not just a lack of cognitive knowledge, but it also reflects an absence of the emotional trauma which makes those events hold more weight. Research on "Depression Babies" shows that people who experienced the 1929 Great Depression didn't actively participate in the market for decades after, even when investing was financially acceptable, because their lived experiences outweighed the market data (Malmendier, 2021). Generation Z has no equivalent formative experience since the most significant event was the COVID-19 market crash of 2020, which was severe but had a sharp V-shape recovery due to major government and central bank intervention (Roberts, 2026).

Rather than instilling caution, the COVID-19 crash made the market seem as if it will always recover (Roberts, 2026). This has created a generation of investors who view risk as an opportunity instead of threat, opposite of Depression Babies. After seeing meme stocks rise just because of hype, even when the company itself isn't actually that valuable, many Gen Z investors feel like traditional ways of analyzing stocks don't really matter as much (Huffman, 2025).

1.2.3 Financial Nihilism

"Financial nihilism" is the growing belief that the current economic system is too broken to build real wealth through traditional means, making conventional investing feel pointless (Roberts, 2026). This mindset is a mix of lived experiences, not just information consumed online. It is driven by the direct experience of rising house prices that make homeownership seem unattainable, the weight of student debt, and anxiety about climate change for some, which makes long-term planning feel pointless (Roberts, 2026; Hillyer & Guseva, 2026; Sullivan, 2026). Within this mindset, taking big financial risks doesn't seem reckless but instead feels logical, since playing it safe seems to guarantee average results anyway. Survey data also supports this, as around 80% of Gen Z believe high-risk investments are a better way to achieve their financial goals than conventional methods (Hardy, 2026). Financial nihilism turns risk-taking from something seen as a bad habit into something that feels like a logical strategy shaped by real economic experiences.

1.2.4 Experiential Knowledge as a Double-Edged Sword

While personal experience is normally the strongest factor which helps stop risky investing behaviors, its effect depends heavily on what kind of experience someone has had. Direct investing experience is the most effective way for forming good financial judgement, more so than formal education alone (Birje et al., 2025). However, if early experiences are overly positive, similar to the 2021 meme stock surges, they may do the opposite and promote risky behavior again by making high-risk investments seem like a reliable strategy which can be misleading (Malmendier et al., 2017).

Gaps in the Literature and Research Agenda

Despite the growing research on Gen Z investment behavior, several questions still remain unanswered.

Gap 1, Research Question 1, and Hypothesis 1

Gap 1: Comparing Cognitive and Experiential Knowledge Across Demographics

Research rarely compares cognitive knowledge and real-life experience within similar samples, often grouping them as "financial literacy" and overlooking demographic differences, leaving it unclear which has a greater impact on behavior (Olschewski et al., 2023).

Research Question 1

How does cognitive versus experiential knowledge of financial crises relate to investment behaviors across demographic groups, including age, gender, and income?

Hypothesis 1

When examining associations across demographic groups, real-life exposure to financial crises is expected to influence post-crisis investment behavior more than cognitive financial knowledge, based on the experience effect and the EBL model.

Gap 2, Research Question 2, and Hypothesis 2

Gap 2: Which Knowledge Type Wins When Both Are Present?

It remains unclear whether learned knowledge or personal experience has a greater influence on investment decisions—for example, Gen Z investors may know about the 2008 crisis but have only experienced a positive market—since no study has directly compared these effects when both are present.

Research Question 2

When an individual has both cognitive knowledge of financial crises and direct market experience, which knowledge type more strongly influences their investment risk behavior?

Hypothesis 2

Greater financial knowledge slightly reduces risk-taking, but real-life experience still has a stronger influence on investment decisions.

Gap 3, Research Question 3, and Hypothesis 3

Gap 3: Do Demographic Factors Change the Cognitive-Experiential Dynamic?

Research rarely studies whether the correlation between knowledge type and investing behavior varies by demographics like age or gender, remaining unclear if these effects apply broadly or differ across groups.

Research Question 3

Does the relative influence of cognitive versus experiential knowledge on investment behavior vary significantly across age groups and genders?

Hypothesis 3

The effect of cognitive versus experiential knowledge on investing behavior is expected to vary across different age groups and genders, with age having the biggest impact since it reflects more market experience.

Gap 4, Research Question 4, and Hypothesis 4

Gap 4: Does Social Media Simulate Experiential Learning?

While it is well known that Gen Z is highly involved in cryptocurrency and meme stocks, it remains unclear whether Gen Z's high-risk investing is caused by low knowledge, financial nihilism, or a logical reaction to structural economic conditions (Roberts, 2026; Hardy, 2026; Mucskova et al., 2025). One idea that hasn't been studied enough is that social media exposure might act as a substitute for real investing experience.

Research Question 4

Does exposure to financial content on social media increase risk-taking behavior by acting as a substitute for real-world investing experience, particularly among younger individuals with limited direct crisis exposure?

Hypothesis 4

Exposure to financial content on social media can increase risk-taking by creating a "fake experience" of investing, especially for younger users who are shown success stories through algorithms, which can make them feel informed and lead to more risky decisions similar to real market experience.

2. Methods

2.1 Participants

An online survey of 163 American adults within matched age and income groups was conducted in February 2026 through SurveyMonkey, with responses post-stratification weighted to match U.S. population distributions by age, income, gender, and Census region. The original sample was drawn from SurveyMonkey Audience and included 163 U.S. adults across all age groups (18–24: n = 3; 25–34: n = 30; 35–44: n = 31; 45–54: n = 26; 55–64: n = 35; 65+: n = 38 weighted units). Respondents earning below $7,500 or above $250,000 were present in the raw data and retained after post-stratification weighting; the income range of $7,500–$250,000 describes the target recruitment frame, not a strict exclusion criterion. No attention-check questions were included.

2.2 Procedure and Design

Post-stratification weighting was conducted using age group and income bracket marginals from the 2024 IPUMS USA extract of the American Community Survey. Each participant was assigned a weight (WT) representing the number of U.S. adults they represent within matched age-income cells. Estimates are therefore only generalizable to age-income combinations present in both the survey sample and the ACS data. In subsequent steps, weights

were further adjusted to match the national gender split of approximately 50.5% female and 49.5% male, as well as the population distribution across all nine U.S. Census regions.

2.3 Research Process

The survey measured investment behavior, risk attitudes, and views of financial knowledge using 18 questions. Questions with Likert scales were scored from 0 to 100 and then grouped into four equal-interval categories so they could be analyzed categorically, and open-ended responses were sorted into thematic categories. To see whether age group was related to each outcome, chi-squared tests of independence were used with a significance level of $\alpha = 0.05$. In this analysis, age group is used as the primary proxy for differences in real-world experience with financial decision-making. Analyses were conducted in R and Excel.

Survey items map onto the two knowledge constructs as follows. Cognitive knowledge is captured by Q17 (formal education as investment preparedness) and Q18 (financial knowledge feels irrelevant), which assess attitudes toward learned financial information. Experiential knowledge is captured by Q6 (personal experience of a major financial crisis), Q8 (influences on risk avoidance, including own past losses), Q9 (experience valued over formal education), and Q11 (transformative investment experience). Behavioral outcomes — Q3 through Q5, Q7, Q10, Q12 through Q16 — are treated as dependent variables rather than knowledge measures. Q15 and Q16, which capture social media exposure, are analyzed separately as a potential proxy for vicarious experiential learning, consistent with Hypothesis 4.

Before interpreting chi-squared results, expected cell counts were examined. The 18–24 age group contained only approximately 3 weighted units, producing expected counts below 5 in several cells. For questions with many response categories (Q4, Q8, Q11), some cells had low expected counts; these results should be interpreted with caution as chi-squared assumptions may not be fully met in those comparisons.

3. Results

3.1 Likert Scale Scoring

Several questions in this survey used a Likert scale ranging from 0 to 100. To make these responses easier to analyze, scores were divided into four categories: 0–25 (low agreement or never), 26–50 (neutral-low), 51–75 (neutral-high), and 76–100 (high agreement or always). This approach allows responses to be compared across age groups using chi-squared tests, even though the original answers were on a continuous scale.

3.2 Demographic Overview

Table 1 summarizes the weighted demographic makeup of the sample across six age groups, including gender composition, predominant income bracket, and key behavioral patterns.

Table 1

Weighted Demographic Breakdown by Age Group

Age Group	Weighted Units (~)	Gender Composition	Primary Income Bracket	Notable Pattern
18–24	~3	100% Female (sample)	$100K–$150K	Low representation; limited experience
25–34	~30	~59% Female	$100K–$150K	Highest social media-driven risk-taking
35–44	~31	Mixed	$150K–$250K	Highest crypto use; peak earners
45–54	~26	Near equal	$150K–$250K	Experience-driven caution; past losses cited
55–64	~35	~61% Male	Mixed	High risk avoidance despite male majority
65+	~38	~55% Female	~65% below $100K	Most conservative; financial & experiential caution

In terms of overall income distribution, about 51% of the weighted sample falls below $100,000, 23% is in the $100,000–$150,000 range, 15% is in the $150,000–$250,000 range, and about 11% earns above $250,000. Across the raw data, individuals earning $150,000–$250,000 were overall more inclined to take risks and more open to trending investments. Those earning below $100,000 were hesitant due to limited financial capacity, while those earning above $250,000 were focused more on wealth preservation.

Out of 17 questions tested, 14 showed statistically significant relationships with age groups. Gender did not show a significant overall result ($p = 0.281$), and household income was also not significant ($p = 0.474$). Age, used here as a proxy for real-world financial experience, is the strongest statistical predictor across the survey. However, gender and income still provide some important context to specific results throughout the study.

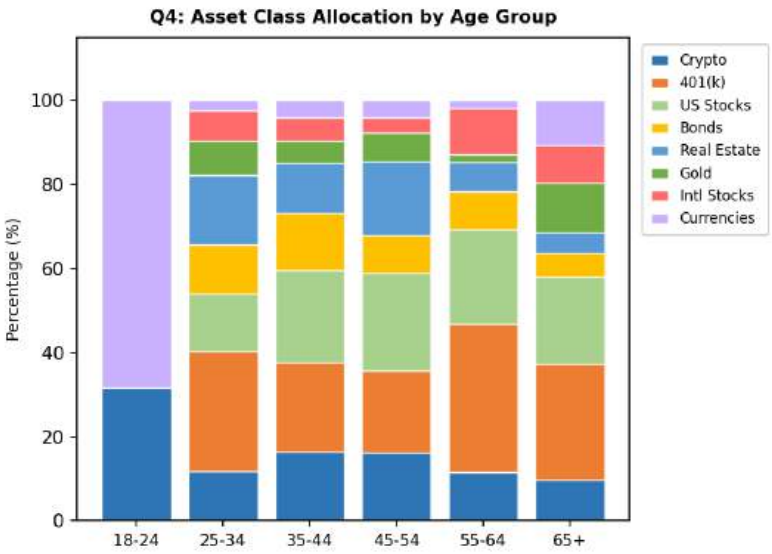


Figure 1. Asset Class Allocation by Age Group (Q4)

Asset class selection showed clear differences across age groups (p = 0.007). The 18–24 group had the largest share of cryptocurrency in their portfolio, as shown in Figure 2, while the 35–44 group had the largest total number of cryptocurrency users when adjusted for group size, which makes sense given that this group had the highest share of earners in the $150,000–$250,000 range — the income bracket most linked to crypto use in this dataset. The 25–34 group also showed strong cryptocurrency participation, which lines up with its large share of earners in the $100,000–$150,000 range. In contrast, the oldest groups held very little cryptocurrency and mostly invested in 401(k)s, bonds, and stocks. Notably, people earning above $250,000 were actually among the most cautious with speculative investments, suggesting that once someone reaches a certain level of wealth, protecting it becomes more important than trying to grow it further.

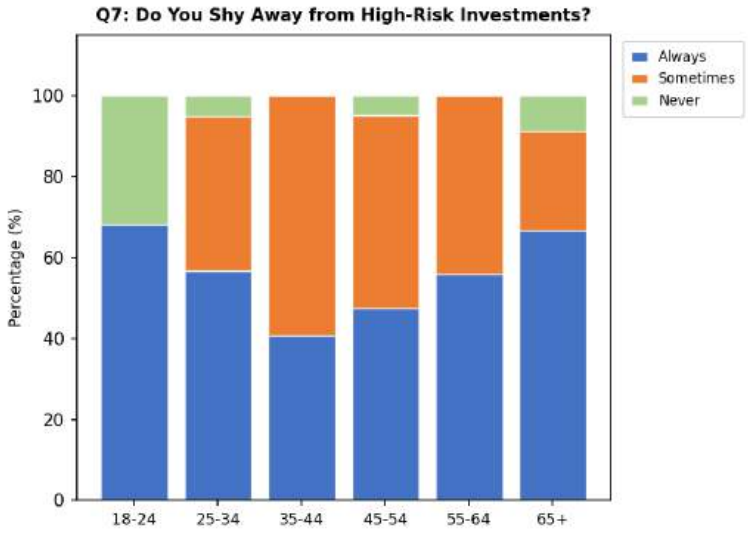


Figure 2. Risk Avoidance Behavior by Age Group (Q7)

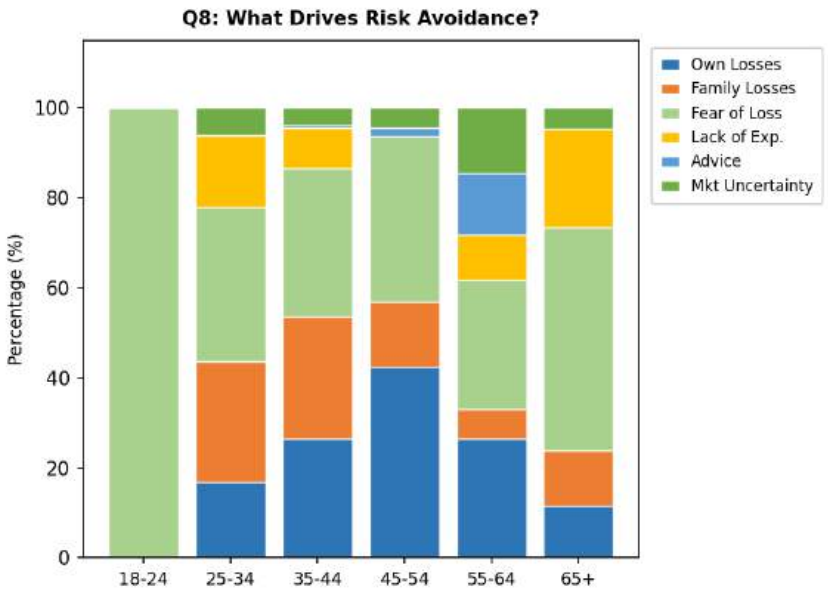


Figure 3. What Drives Risk Avoidance by Age Group (Q8)

Risk avoidance was significantly linked to age ($p = 0.041$). The 65+ group had the highest share of respondents who said they always avoid high-risk investments, a pattern supported by the fact that 65% of this group earns below $100,000, making risky investing more financially difficult rather than just a personal choice. The 55–64 group also showed high risk avoidance even though it was 61% male, a group usually linked with higher risk tolerance, which suggests that age and experience can matter more than gender patterns. The 45–54 group most often pointed to their own past financial losses as the reason for being cautious ($p = 0.010$), showing experience-based avoidance. In contrast, the 25–34 group most often cited a lack of experience, meaning their caution is more about not knowing enough rather than having gone through losses themselves.

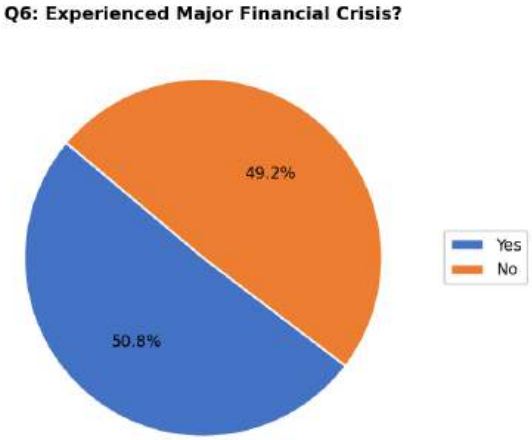


Figure 4. Self-Reported Experience of Major Financial Crisis (Q6)

Figure 4 shows the overall split for Q6 (50.8% yes, 49.2% no); this question was not significantly associated with age group (χ^2=8.96, p=0.111). The findings on risk avoidance described in the previous section relate to Q7 and Q8, not Q6.

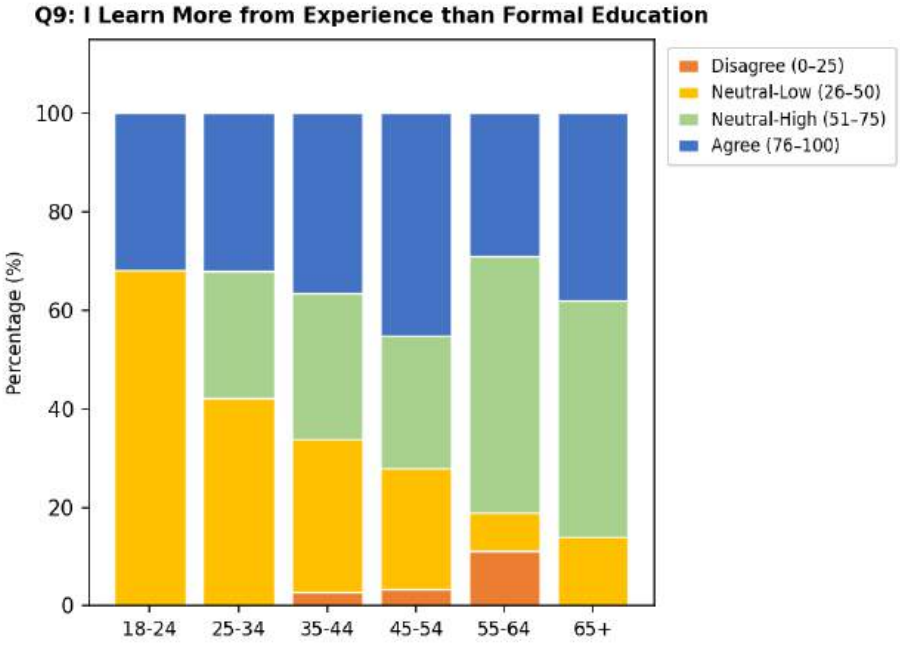


Figure 5. Experience vs. Formal Education as a Learning Source (Q9)

The 65+ and 55–64 groups showed the highest agreement with the idea that personal experience matters more than formal financial education (p = 0.015). The 65+ group has had many years to apply what they learned in real markets, and with 65% earning below $100,000, many have also experienced financial hardship firsthand. The 55–64 group, which was mostly male and included a range of incomes, also strongly agreed that experience matters more, showing that later in life, real-world experience often outweighs formal education regardless of gender or income.

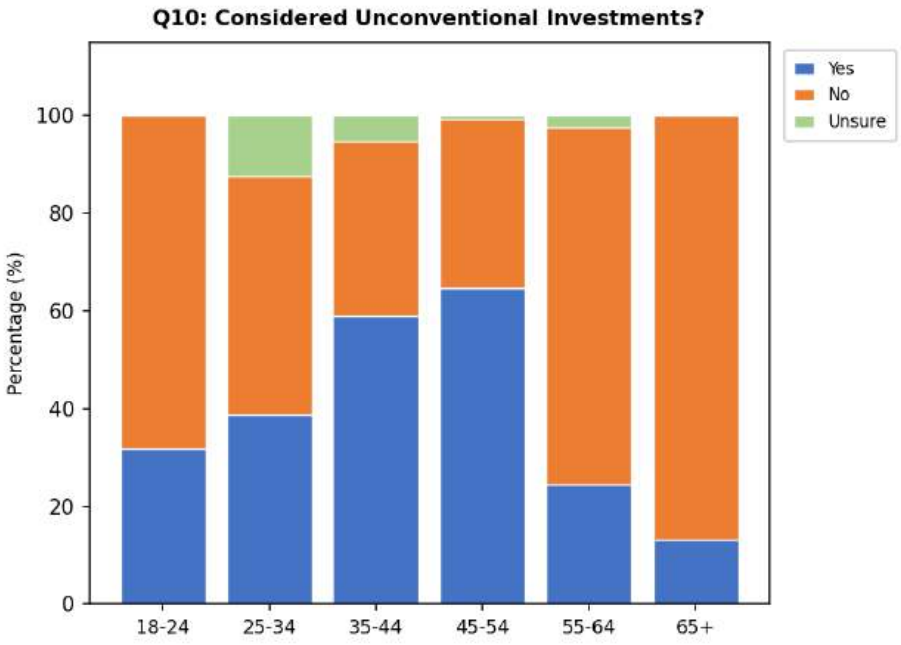


Figure 6. Consideration of Unconventional Investments (Q10)

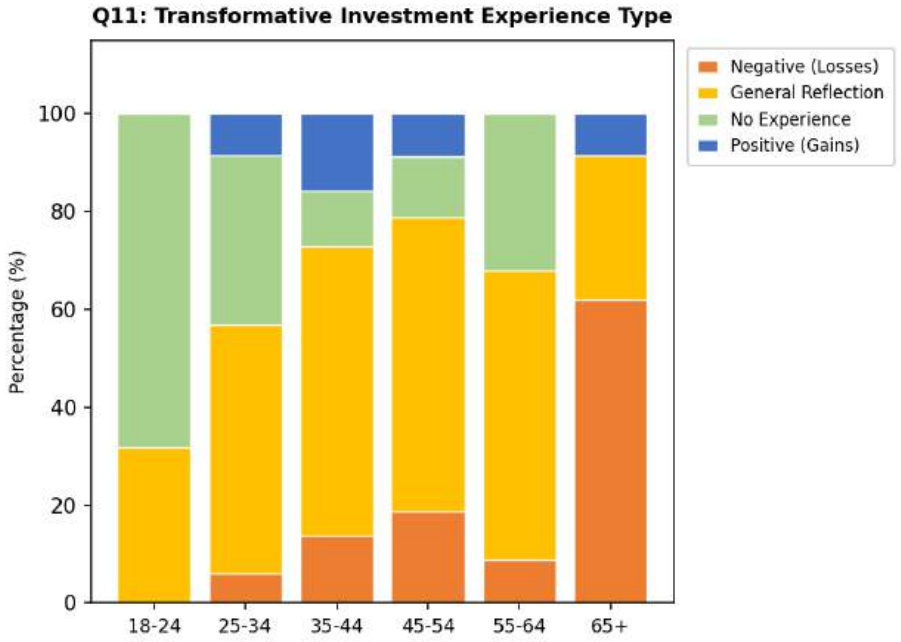


Figure 7. Nature of Transformative Investment Experience (Q11)

Considering unconventional investments was strongly linked to age ($p < 0.001$). The 35–44 group had the highest share of respondents saying yes, consistent with their income profile—individuals earning $150,000–$250,000 were the most open to cryptocurrency, and this age group had the largest proportion of such earners. The 65+ group was almost entirely in the no category, with income and gender both contributing to this conservatism. The question about

transformative financial experiences produced the strongest result in the entire study ($p < 0.001$, $\chi^2 = 62.47$). The 65+ group accounted for nearly 24 weighted units reporting negative experiences such as losses, crashes, and bankruptcies. The 55–64 group focused more on general reflection, suggesting that older male investors with higher incomes tend to process market experiences analytically. Younger groups described investing in abstract terms with little focus on negative outcomes.

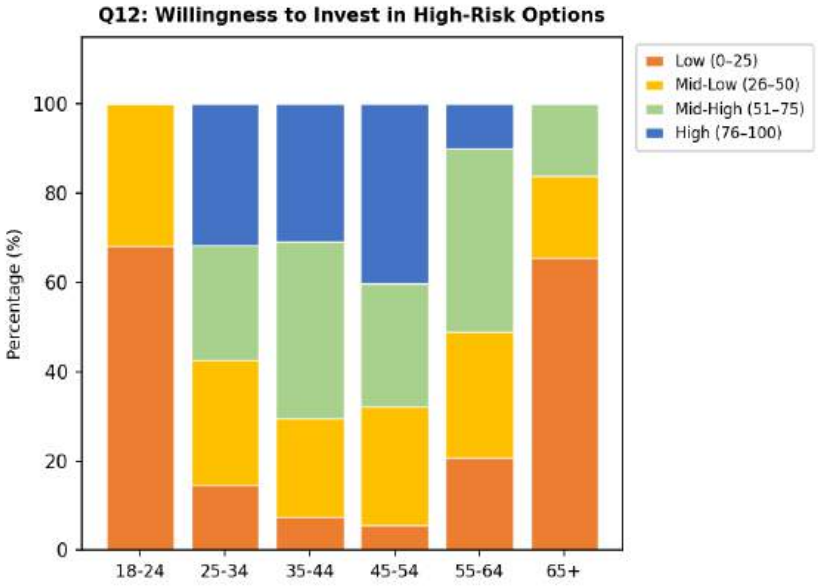


Figure 8. Willingness to Accept High-Risk Investments (Q12)

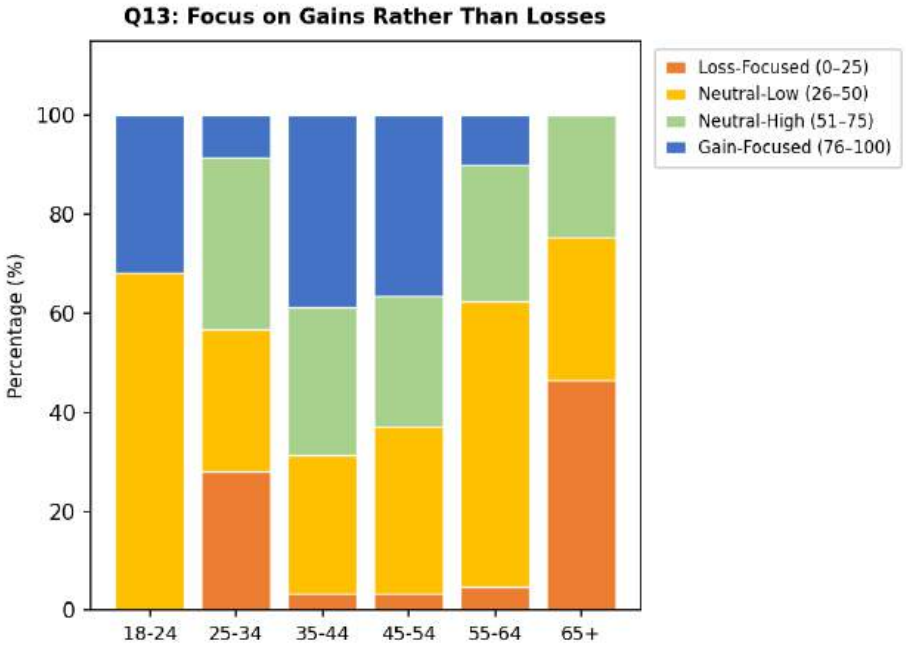


Figure 9. Gain-Focused Decision Making (Q13)

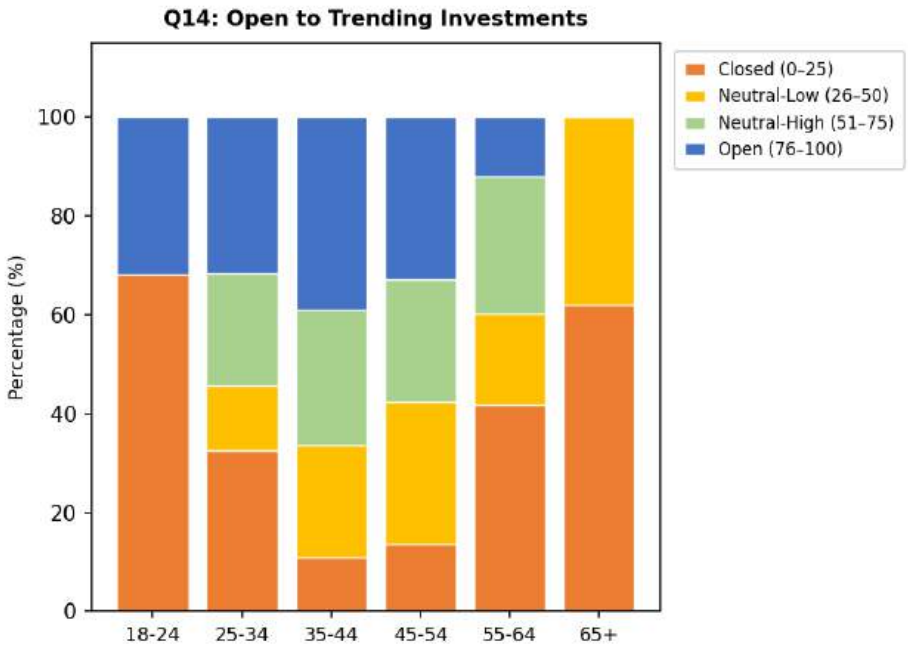


Figure 10. *Openness to Trending Investments (Q14)*

Willingness to take high risks, focus on gains, and openness to trending investments all showed a strong connection to age ($p < 0.001$ for all three). The 35–44 group's greater risk-taking and the 65+ group's higher likelihood to conserve money are both partly explained by income, which is discussed further in the Discussion section. Men were slightly more willing to take risks than women across income levels, but this difference was much smaller than the combined effect of age and income.

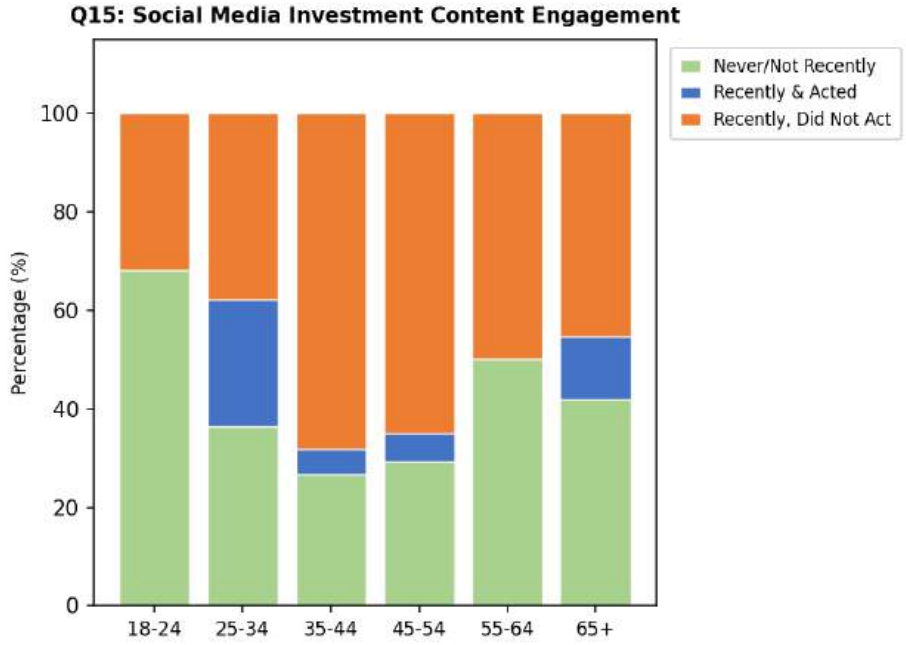


Figure 11. Social Media Engagement with Investment Content (Q15)

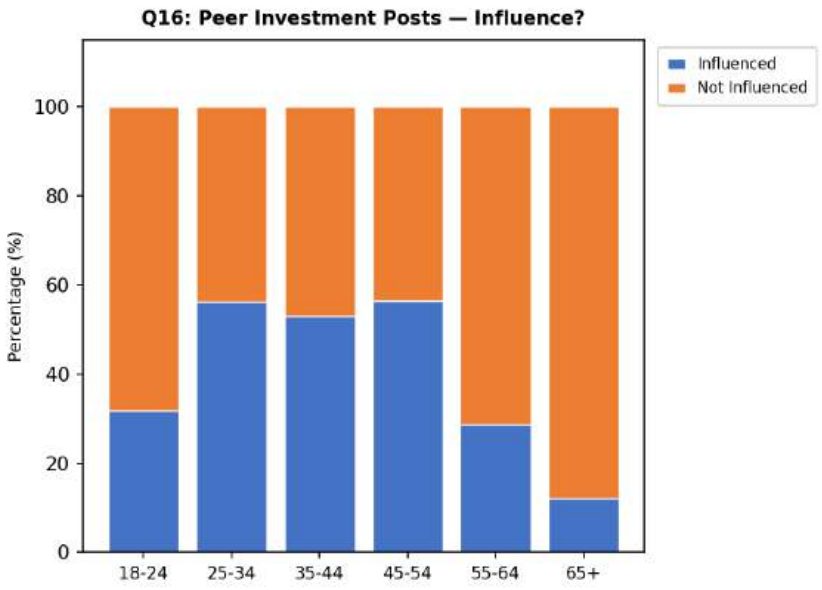


Figure 12. Peer Investment Posts and Their Influence (Q16)

Both social media engagement and peer influence showed a strong link to age ($p = 0.025$ and $p = 0.001$). The 25–34 group, which was 59% female and mostly earned between $100,000 and $150,000, had the highest rates of watching investment content online and acting on it. For investors at this income level with limited market experience, high-risk digital assets may seem like the quickest way to grow wealth, and social media reinforces this by constantly showing success stories. In contrast, the 55–64 group had no respondents who recently acted on social media investment content, and the 65+ group was also largely unaffected. This pattern is consistent with the idea that real-world experience is associated with older investors being less influenced by outside sources.

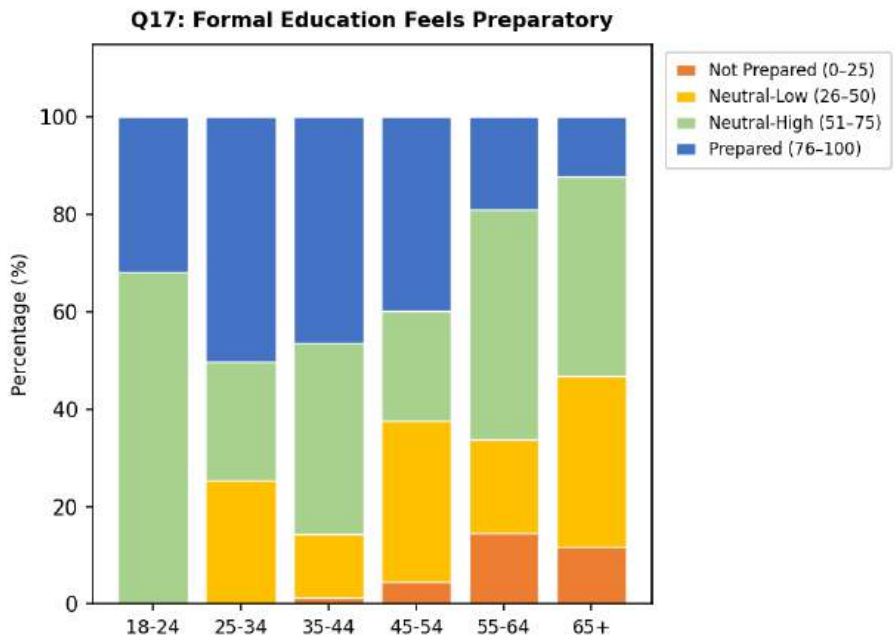


Figure 13. Formal Education as Investment Preparedness (Q17)

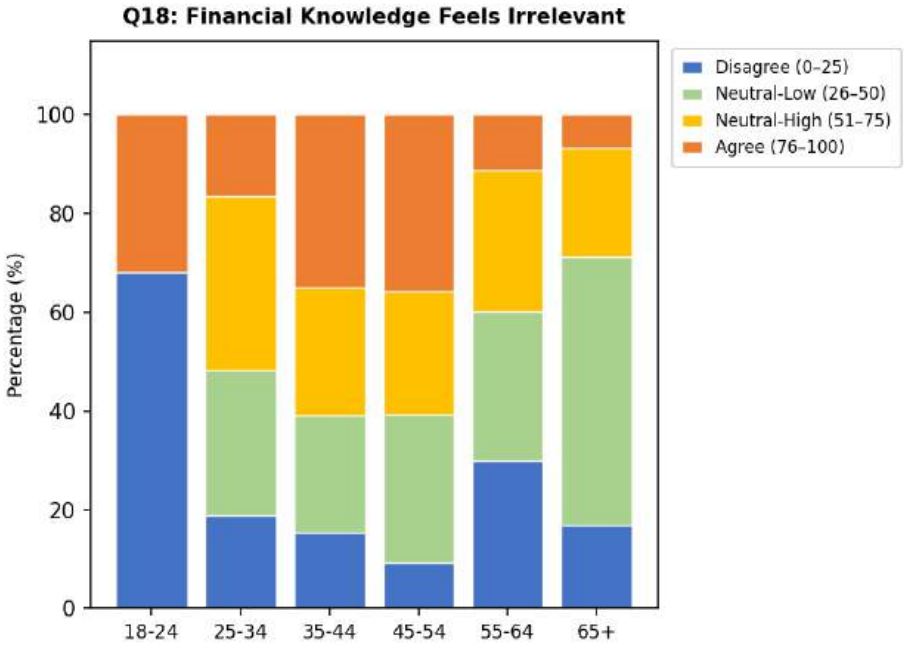


***Figure 14**. Financial Knowledge Feels Irrelevant in Today's Market (Q18)*

Confidence in formal financial education was highest in the 25–34 group and varied significantly across age groups ($p = 0.010$). Men showed slightly higher confidence in formal education overall, and the 35–44 group, which had more male and higher-income respondents, also felt more prepared. The belief that financial knowledge feels less relevant in today's market was strongest in the 35–44 and 45–54 groups ($p = 0.019$), which had wide income ranges and

more exposure to markets where traditional rules did not always seem to apply. Rather than showing a lack of knowledge, this skepticism seems to come from experience and what people have seen in real markets.

Table 2

Chi-Squared Test Summary — All Questions

Question	Topic	χ^2	df	p-value	Significant
Q2	Gender	6.26	5	0.281	No
Q3	Invest / Plan to Invest	22.69	5	< 0.001	Yes
Q4	Asset Class Selection	58.73	35	0.007	Yes
Q5	Current Household Income	14.70	15	0.474	No
Q6	Experienced Financial Crisis	8.96	5	0.111	No
Q7	Shy Away from High Risk	18.96	10	0.041	Yes
Q8	Influences on Risk Avoidance	44.39	25	0.010	Yes
Q9	Experience > Formal Education	29.22	15	0.015	Yes
Q10	Unconventional Investments	35.24	10	< 0.001	Yes
Q11	Transformative Experience	62.47	15	< 0.001	Yes
Q12	Willing: High-Risk / High-Return	61.26	15	< 0.001	Yes
Q13	Focus on Gains > Losses	60.62	15	< 0.001	Yes
Q14	Open to Trending Investments	51.90	15	< 0.001	Yes
Q15	Social Media: Watched & Acted	20.50	10	0.025	Yes
Q16	Peer Posts Influence Behavior	21.98	5	0.001	Yes
Q17	Formal Education = Preparedness	30.52	15	0.010	Yes
Q18	Financial Knowledge Feels Irrelevant	28.43	15	0.019	Yes

When the chi-squared results are considered along with all three demographic factors, a clear pattern appears. Age is the strongest predictor in 14 of the 17 questions, while gender and income do not show strong independent links with age group. However, they still help explain the size and direction of the patterns seen in the age data. The strong caution in the 65+ group is supported by a higher share of women and its income level, since 65% earning below $100,000 suggests many people may not have enough financial capacity, in addition to experience-based caution. The higher risk-taking in the 35–44 group is supported by its large share of $150,000–$250,000 earners, the income group most linked with risk-taking in this dataset. The strong chi-squared values on risk questions show these patterns are not random but reflect real,

population-weighted differences in behavior. These differences are most strongly associated with age, with income and gender playing smaller supporting roles. While the pattern is consistent with an experiential knowledge explanation, the non-significance of Q6 means real-world crisis exposure cannot be confirmed as the direct mechanism.

4. Discussion

When age, gender, and income are all examined alongside one another, the data show a more complete picture. The main finding is still the same, experiential knowledge is the main predictor of investment behavior. However, income and gender help explain why different groups act the way they do. The strong caution seen in the 65+ group is not only correlated with their market experience. 65% of this group earns below $100,000, meaning many rely on fixed or limited retirement incomes where any financial loss can have serious effects. For these respondents, avoiding speculative investments is both a personal choice and a financial necessity. Similarly, the 35–44 group's high use of cryptocurrency reflects not only experience-based optimism but also the fact that this group has the largest share of $150,000–$250,000 earners who can realistically afford to handle possible losses.

One interesting income-related finding is that respondents earning above $250,000 were actually the most conservative out of all income groups, even more than those earning $150,000–$250,000. This focus on protecting wealth goes against the idea that more money automatically leads to more risk-taking. It suggests that after reaching a certain level of wealth, keeping what someone has becomes more important than trying to grow it. This matches earlier research showing that experienced, wealthy investors tend to shift from focusing on growth to focusing on protecting their money as their financial situation becomes more stable.

Gender adds some detail to the findings without changing the main results. Men were slightly more willing to take risks and focus on gains than women, with about a six-point difference on the 0–100 risk scale. What is more important is how gender works with age. The 55–64 group is 61% male, which might suggest higher risk-taking, but it does not, showing that age and experience matter more than gender. The 25–34 group is 59% female, which might suggest lower risk-taking, but it also does not, because the lack of negative real-world experience seems to matter more than gender-based patterns. This connects to the description-experience gap, where younger investors may understand risk in theory but have not actually experienced it in a way that changes their behavior.

The social media findings become more important when income is also considered. The 25–34 group, which is 59% female and mostly earns $100,000–$150,000, shows the highest levels of both social media investment engagement and cryptocurrency use. At this income level, and with limited market experience, high-risk digital assets may seem like the most realistic way to build wealth quickly. Social media reinforces this by constantly showing stories of big financial gains. In contrast, older investors in the 55–64 and 65+ groups, who are shaped more by experience and often limited by income, are mostly not influenced by these platforms in the same way.

Moving on to the hypotheses, hypothesis 1 predicted that experiential knowledge would be a stronger predictor than cognitive knowledge. This is partially supported: age, used as a proxy for experiential knowledge, was the strongest predictor across 14 of 17 questions. However, Q6, the only direct measure of personal crisis experience, was not significant ($p = 0.111$), which limits how strongly this hypothesis can be confirmed. The income and gender data also support this by showing that the 65+ group's caution comes from both past market experience and financial limits. Hypothesis 2, which predicted that cognitive knowledge would reduce but not fully stop risky investing, is partially supported. The 25–34 group is confident in formal financial education but still has high cryptocurrency use and social media engagement, showing that knowledge alone does not fully change behavior. Hypothesis 3 predicted differences based on age and gender. Age is clearly the most important factor across the results, while gender is not significant overall but still helps explain smaller patterns within groups. Hypothesis 4, which suggested that social media can act like a substitute for real experience, is supported. The social media effect is strongest in younger groups no matter their incoy.

4.1 Limitations

There were several limitations in the study. To begin with, chi-squared tests look for correlations, but cannot necessarily establish any form of causation. The original survey's income limits were partly adjusted using ACS population weighting, which makes the results more accurate but also harder to understand. Gender was only counted as two options because of the data, which leaves out more detail.

4.2 Reflections and Addressing the Research Gap

This study addresses all four gaps in the research. By using population weighted data from six age groups, it shows that experiential knowledge is the strongest factor in how people invest even after considering income and gender. The results also demonstrate that, when directly compared, real life experience holds more weight than reading about a crisis. The income results add another layer by suggesting that some younger investors with medium incomes may take risks on purpose instead of out of confusion. The social media results also address the last gap by showing that online content can act like secondhand experience and shape how people think about risk, even if they have not experienced a market crash themselves.

4.3 Implications

These findings show that current financial literacy programs are insufficient because cognitive knowledge does not hold the same weight as the emotional impact of personally experiencing market crises. Current programs should focus on providing that real-world experience many youth currently lack by simulating real-world market scenarios.

4.4 Future Directions

Future research should examine more variables to better understand how investment behavior is shaped. This studyme or gender, which suggests it is mostly linked to lack of real market experience.

Regarding the literature gaps, the first gap called for comparisons of cognitive and experiential knowledge within similar samples across demographics. This study addresses that by showing that experiential knowledge is still the strongest predictor even after accounting for age, gender, and income. The second gap asked which type of knowledge matters more when both are present. The data show that even a small amount of real-life experience often matters more than formal education when explaining risky investing behavior. The third gap examined whether the relationship between knowledge and behavior changes across demographics. Age had a bigger effect, while gender was not as strong. The income results also help explain the fourth gap, which is about structural reasons. For investors earning $100,000–$150,000, cryptocurrency may seem like a reasonable choice because traditional investments may not help them build wealth fast enough to close the gap they face.

5. Conclusion

This study set out to answer whether simply knowing about financial crises is enough to make someone invest more carefully, or whether a person has to actually live through a market crash to develop that caution. The results clearly showed that age was the strongest predictor of investment behavior, doing a better job of explaining differences than formal education, gender, or income. Older investors were significantly more likely to avoid high-risk investments (Q7, $p=0.041$) and more often pointed to their own past financial losses as the reason for that caution (Q8, $p=0.010$), which lines up with the idea that real-life experience shapes behavior. However, Q6, which directly asked whether someone had personally experienced a major financial crisis, was not significant ($\chi^2=8.96$, $p=0.111$), meaning that crisis experience alone cannot be confirmed as the cause — the pattern seems to be tied to age more broadly rather than any single experience. Younger investors, who mostly came of age during the fast 2020 market recovery, were more willing to take risks, more influenced by social media, and more open to trending investments like cryptocurrenc focuses mainly on age, gender, and income level; however, other variables such as ethnicity and region can also play an important role. Specifically, analyzing ethnicity can show how different cultural backgrounds shapes risk tolerance, while studying different regions can show how local factors change how people save and spend.

Works Cited

Ahmmad, M., Shahzad, K., Iqbal, A., & Latif, M. (2025). Trap of Social Media Algorithms: A Systematic Review of Research on Filter Bubbles, Echo Chambers, and Their Impact on Youth. Societies, 15(11), 301. https://doi.org/10.3390/soc15110301

Anggiani, L., Laksmana, K., & Ardani, W. (2025, September 28). Investment in the Digital Era: The Influence of Digital Social Influence and Behavioral Finance on Generation Z's Investment Decisions | Formosa Journal of Multidisciplinary Research. Npaformosapublisher.org. https://npaformosapublisher.org/index.php/fjmr/article/view/507/575

Asri, S., Paulina, J., Tarigan, B. A., & Siagian, A. O. (2024). Role of Financial Literacy, Risk Perception, and Investment Experience in Investment Decisions of Millenial and Z Generation in Jabodetabek. International Journal of Business, Law, and Education, 5(2), 1750–1761. https://doi.org/10.56442/ijble.v5i2.731

Birje, R., Kulkarni, S., M, M., & K, V. (2025). Exploring the Determinants of Investment Decisions among Generation Z: The Role of Financial Education, Experience, and Risk Tolerance. Unep.org, 2(5), 310–318. https://doi.org/N/A

CFA Institute. (2023). Gen Z and Investing: Social Media, Crypto, FOMO, and Family. In www.finrafoundation.com. https://rpc.cfainstitute.org/sites/default/files/-/media/documents/article/industry-research/Gen_Z_and_Investing.pdf

Darwish, F. (2025). Financial literacy and investment decision: an empirical study from the Palestine Stock Exchange. Frontiers in Behavioral Economics, 4. https://doi.org/10.3389/frbhe.2025.1444022

Dhamma, I., & Setyawati, C. Y. (2026). How Financial Literacy Influences Gen Z Investment Decisions through Attitude and Overconfidence. Edu Cendikia: Jurnal Ilmiah Kependidikan, 5(03), 1412–1424. https://doi.org/10.47709/educendikia.v5i03.7667

Hardy, A. (2026, March 10). "Financial Nihilism" Is Pushing Young Americans Toward Crypto, Sports Betting and More (K. Peach, Ed.). Money. https://money.com/financial-nihilism-gen-z/

Hii, I., & Ong, Y. (2025). Finfluencer: can financial social media influencers promote desirable financial behaviours? International Journal of Bank Marketing, 44(2). https://doi.org/10.1108/IJBM-05-2024-0256

Hillyer, M., & Guseva, N. (2026, January 14). Why does Gen Z invest so differently? It's a trust issue. World Economic Forum. https://www.weforum.org/stories/2026/01/gen-z-financial-investment-trust/

HRB Tax Group, Inc. (2025, March). A Survey by Spruce Reveals Social Media's Growing Influence on Gen Z's Financial Decisions, Highlighting a Generational Divide in Learning about Money. Nasdaq.com; H&R Block. https://www.nasdaq.com/press-release/survey-spruce-reveals-social-medias-growing-influence-gen-zs-financial-decisions

Huffman, N. (2025). Rise of Meme Stocks: How Gen Z Investors Influence Market Volatility. Social Science Research Network. https://doi.org/10.2139/ssrn.5514739

Malmendier, U. (2021). Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3935524

Malmendier, U., Pouzo, D., & Vanasco, V. (2017). A Theory of Experience Effects *. https://eml.berkeley.edu/~ulrike/Papers/MPV_2017-01-15.pdf

Mucskova, M., Micak, P., & Nemecek, K. (2025). Generation Z and Financial Literacy: A Comparative Analysis of Generational Competencies. Liberec Economic Forum 2025, 149–160. https://doi.org/10.15240/tul/009/lef-2025-12

Olschewski, S., Luckman, A., Mason, A., Elliot Andrew Ludvig, & Konstantinidis, E. (2023). The Future of Decisions From Experience: Connecting Real-World Decision Problems to Cognitive Processes. Perspectives on Psychological Science, 19(1). https://doi.org/10.1177/17456916231179138

Price, D. A. (2024). Ulrike Malmendier. Richmondfed.org; Econ Focus. https://www.richmondfed.org/publications/research/econ_focus/2024/q1_q2_interview

Roberts, L. (2026, February 13). Financial Nihilism & The Trap Young Investors Are Walking Into. Realinvestmentadvice.com. https://realinvestmentadvice.com/resources/blog/financial-nihilism-the-trap-young-investors-are-walking-into/

Sheela Sundarasen, Kamaludin, K., & Ibrahim, I. (2026). From Family Systems to Financial Outcomes: Role of Parental Financial Socialization. Societies, 16(2), 47–47. https://doi.org/10.3390/soc16020047

Spohn, D. (2024). Financial Resilience and Innovation among Generation Z in the Face of Economic Adversity. European Journal of Management, Economics and Business, 1(3), 39–51. https://doi.org/10.59324/ejmeb.2024.1(3).04

Sullivan, L. (2026, January 22). Upticks: Is Gen Z Giving Up on Money? Falcon Wealth Advisors. https://falconwealthadvisors.com/blogs/upticks/upticks-is-gen-z-giving-up-on-money

The Echo. (2026, March 17). AI, social media influencing Gen Z's financial decisions. The Echo. https://www.echo.net.au/2026/03/ai-social-media-influencing-gen-zs-financial-decisions/

Field, A. (2024). Discovering statistics using IBM SPSS statistics (6th ed.). SAGE Publications.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 1–55.

Lumley, T. (2010). Complex surveys: A guide to analysis using R. Wiley.

R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., Richards, S., Rodgers, R., & Schouweiler, M. (2024). IPUMS USA: Version 15.0 [Dataset]. IPUMS. https://doi.org/10.18128/D010.V15.0

The Effect of Global Financial Crisis and COVID-19 Pandemic on Return on Assets: Evidence From Stock Exchange of Thailand By Torfan Wantarnagon

Abstract

The paper examines the effect of the Global Financial Crisis (GFC) and the COVID-19 pandemic on profitability, measured by Return on Assets (ROA), across eight sectors listed on the Stock Exchange of Thailand (SET) from 2001 to 2024. The study uses regression analysis and extensive descriptive statistics to evaluate the impact of these two major crises on firm performance. The results show that both the GFC and COVID-19 pandemic had a statistically significant negative effect on ROA. However, the COVID-19 pandemic had a much larger impact on corporate profitability than the GFC. The sectoral analysis also demonstrates significant differences in resilience and recovery rates across different industries. The Agriculture and Food Industry (AGRO) sector was the most resilient during both crises, while the Property and Construction (PROPCON) sector experienced the largest decline in profitability and the weakest recovery among the eight sectors. These results suggest financial and public health crises can significantly reduce corporate profitability and highlight the importance of sector-specific policies to improve resilience during periods of systemic risk.

Keywords Return on Assets (ROA), Global Financial Crisis (GFC), COVID-19, Stock Exchange of Thailand, regression analysis

Introduction

The stability and profitability of firms are often affected by major global events such as the Global Financial Crisis (GFC) and the COVID-19 pandemic. Understanding how these crises influence firm performance is important for investors, policy makers, and business managers, particularly in emerging markets.

The purpose of this paper is to examine how the Global Financial Crisis (GFC) and the COVID-19 pandemic affected corporate profitability, measured by Return on Assets (ROA), across eight different sectors listed on the Stock Exchange of Thailand (SET) between 2001 and 2024.

Profitability patterns changed during the Global Financial Crisis (GFC) of 2008-2009. Prior to the crisis, many firms benefited from rising asset prices and strong external demand which helped support a higher average ROA. During the crisis, falling asset values and weaker trade reduced sales and increased overall risk, thereby causing profitability to decline. In the post-crisis period, average ROA generally gradually improved but not evenly across all sectors or countries. Studying profitability before, during, and after the GFC is crucial as it demonstrates which sectors are most vulnerable to financial crises and how quickly firms recover after a major disruption.

Profitability was also disrupted significantly by the COVID-19 pandemic. Before the pandemic, many firms in both developed and emerging markets displayed stable or gradually

increasing average ROA. During the COVID-19 pandemic, lockdowns, supply chain disruptions, and demand shocks generally lowered revenue and increased costs. In combination, this caused profitability to decrease in various sectors. However, some essential and technology-related sectors were more resilient to the crisis. After the pandemic, average ROA partially recovered at different rates across sectors. Examining profitability before, during, and after the COVID-19 pandemic highlights which sectors are most vulnerable to public health crises and how firms adapt in response.

The main findings of this study indicated that both the Global Financial Crisis (GFC) and the COVID-19 pandemic had a statistically significant negative impact on ROA. The regression analysis indicated that the coefficients of the GFC and COVID-19 pandemic were -0.893 (significant at the 10% level) and -1.295 (significant at the 1% level) respectively. This suggests that the COVID-19 pandemic had a more severe effect on Thai corporate performance than the GFC.

The paper contributes to the existing literature by providing a long-term comparative analysis for two different types of global crises–financial and public-health-related–within the context of an emerging market. Unlike previous studies that often focus on a single event or industry, the present paper focuses on a broader sector perspective covering the agriculture & food, consumer products, financials, industrials, property & construction, resources, services, and technology sectors listed on the Stock Exchange of Thailand.

The remainder of the paper is organized as follows. The next section reviews the literature, followed by the empirical methodology, results, and conclusion.

Literature Review

Firm performance during the period of economic and financial instability has attracted significant attention in the literature. The effect of global shocks including financial crises on profitability, liquidity, investment decisions, and overall corporate finance has been thoroughly investigated. Return on Assets (ROA) is one of the most popular indicators of profitability.

There is a large amount of literature on the effects of global financial crises on firm performance. For example, Lai et al. (2014) found that Malaysian firms' profitability declined during the crisis period and the revenue growth was sustained through government aid. Similarly, Yap et al. (2014) found that the financial performance of Malaysian firms had substantially declined after the crisis. In Australia, Thangaraj and Chan (2012) indicated that the crisis affected different industries differently. Nazir et al. (2012) found that asset quality was the strongest factor impacting ROA in Pakistan during the financial crisis. Olaniyi and Olabisi (2011) found that despite high liquidity, Nigerian banks were negatively impacted by the crisis. Kočišová (2014) identified capitalization, liquidity, credit quality, operating efficiency, and market structure as the key determinants affecting bank profitability in V4 countries during the crisis. Dietrich and Wanzenried (2011) found that profitability was mainly affected by operational efficiency, loan growth, funding costs, business model, and tax rate. They also found that, during the crisis, Swiss state-owned banks were relatively more profitable. Overall, these

papers show that financial crises affect business profitability and influence financial performance indicators.

Similar findings were observed during the COVID-19 pandemic. According to Shen et al. (2020), the pandemic had a negative effect on firm performance in China. The negative effect was stronger in heavily affected regions and could be lessened with more investment and revenue. Zhang et al. (2024) found that supply chain issues, labor shortages, and financial problems were the main factors that substantially decreased UK construction firm performance during the pandemic. Likewise, Achim et al. (2022) found the Romanian firms' net profits decreased by 37.43% during the pandemic. Alsharef et al. (2021) also found that U.S. construction firms experienced delays, supply problems, lower productivity, and rising material prices during the COVID-19 pandemic period. Overall, these studies suggest that the COVID-19 pandemic has a negative effect on operational and financial performance of various companies for many different regions and sectors.

Several studies have examined business resilience and recovery after major crises. For example, Prasad et al. (2015) found that Korean, Indonesian, and Thai firms had different strengths and speeds at which profitability recovered. Bris et al. (2004) also found that Asian firms became more fragile and less profitable after crises while European and Latin American firms showed clear recovery.

Various other studies further examine firm performance and resilience during crises. For example, Horta et al. (2012) and Peric and Vitezic (2016) show that macroeconomic conditions and firm size or specialization strongly affect construction and industrial performance in Portugal and Croatia during downturns. Lee et al. (2017) and Batool and Sahi (2019) find that ownership concentration, firm age, size, liquidity, leverage, and macro variables such as oil prices and inflation significantly influence crisis-period profitability in China, the United States, and the United Kingdom. Sufian and Habibullah (2010) and Song and Lee (2012) highlight that banks and firms in East Asia adjusted diversification, capitalization, and cash-holding policies after the Asian financial crisis, affecting their post-crisis profitability. Phan and Narayan (2020) emphasize that global stock markets initially overreacted to early COVID-19 news and later partially corrected as more information became available. Farhat and Hili (2021) found that Islamic-compliant stocks in the United States performed better than stocks that were not Islamic-compliant during the COVID-19 pandemic, showing lower volatility rather than higher risk.

Previous research on the financial crisis or the COVID-19 pandemic provided useful information but most focused on one event, industry, or developed economy. Few studies compare both crises, especially in emerging markets. Few studies have examined sectoral profitability differences over a long period of time. Thus, this paper examines how the Global Financial Crisis and COVID-19 pandemic affected the Return on Assets of eight important sectors from the Stock Exchange of Thailand over the period of 2001-2024.

Table 1. Summary of Literature Review

Author	Year	Key variables	Method	Key findings	Region	Time Period
Lai et al.	2014	Current ratio, P/E ratio, Altman's Z-score	Financial analysis	The crisis cut profits, but revenues grew from government aid. Firms cut costs and debt, recovering to 5% profit by 2010	Malaysia	2005-2010
Zhang et al.	2024	ROA, turnover growth, working capital	Regression analysis	COVID-19 reduced UK construction performance in 2020 due to supply, labor, and financial issues, followed by partial recovery in 2021.	United Kingdom	2015-2021
Thangaraj and Chan	2012	Revenue, net assets, quick ratio	Distress analysis	The GFC mildly affected Australian suppliers and contractors but caused large losses in 2009 for developers and REITs.	Australia	2005-2010
Yap et al.	2014	Cash to Current Liability, inventory turnover, cash flow to total assets	Factor and cluster analysis	Malaysian manufacturing firms saw strong performance declines after the GFC, with many previously strong firms shifting into weaker performance groups.	Malaysia	2006-2010
Achim et al.	2022	ROE, net profits, sales, indebtedness, liquidity, working capital, sector	Panel data analysis	Net profits fell 37.43% during the pandemic period. Equity financing, better liquidity management, and larger size supported stronger performance.	Romania	2019 - 2020
Farhat and Hili	2021	Compliance status, abnormal returns, firm-specific volatility	Difference-in-difference regressions	Islamic-compliant stocks outperformed non-compliant stocks during COVID-19, with lower volatility rather than higher risk.	United States	2020
Alsharef et al.	2021	Delays, material shortages, productivity, price escalation, safety measures	Qualitative interviews	COVID-19 caused delays, material shortages, reduced productivity, and rising costs in U.S. construction, prompting safety measures and risk management responses.	United States	2020
Horta et al.	2012	Financial autonomy, value added, size, specialization	DEA	Construction performance improved initially but slowed later, with firm size, specialization, and macroeconomic conditions influencing overall performance differences.	Portugal	1996 - 2009
Shen et al.	2020	Investment growth, revenue, leverage, region and industry impact	Forecasting and robustness tests	COVID-19 reduced firm performance in China, especially in severely affected regions, while higher investment and revenue helped reduce negative impacts.	China	2013 - 2020
Peric and Vitezic	2016	Firm size, turnover growth, sector	Dynamic panel data model	During the recession, turnover growth was positively related to firm size. Medium and large firms grew faster than small firms.	Croatia	2008 - 2013
Lee et al.	2017	Firm age, firm size, leverage	Panel data analysis	During the global financial crisis, firms with higher ownership concentration and older firms performed better, while larger firms performed worse.	China	2008
Phan and Narayan	2020	Country responses, lockdowns, travel bans	Commentary	Stock markets appeared to overreact early to COVID-19 news and later correct as more information became available.	Global	Early 2020

Nazir et al.	2012	Solvency, advances, investments,	Stepwise multiple regression	For Pakistani banks, asset quality was the strongest determinant of bank ROA, followed by size and solvency.	Pakistan	2008
Olaniyi and Olabisi	2011	Loans and advances, customer deposits, investment in securities	OLS multiple regression	The global financial crisis negatively affected Nigerian banks despite high liquidity after consolidation.	Nigeria	2004 - 2009
Kočišová	2014	Bank profitability, capitalization, credit quality	Panel data	Profitability differences among V4 banks were mainly explained by capitalization, liquidity, credit quality, operating efficiency, and market structure.	V4 countries	2007 - 2009
Dietrich and Wanzenried	2011	Operational efficiency, loan growth, funding cost	Dynamic panel GMM	Bank profitability depended on operational efficiency, loan growth, and funding costs, with state-owned banks performing better during crisis periods	Switzerland	1999 - 2009; 2007-2009
Batool and Sahi	2019	Ownership concentration, leverage, liquidity	Panel data	Firm performance was influenced by internal financial factors and macroeconomic conditions, with differing effects observed between U.S. and U.K. markets.	United States and United Kingdom	2007 - 2016
Prasad et al.	2015	Gross profit margin, net profit margin, basic earning power	Ratio analysis	Profitability recovery after the crisis differed across Indonesia, Korea, and Thailand. Recovery speed and strength were not the same across countries.	Indonesia, Korea, Thailand	1994 - 2004
Bris et al.	2004	Leverage, financial fragility, exchange-rate exposure	Descriptive analysis	Financial crises reduced profitability and increased fragility, especially in Asia, while recovery patterns differed, with clearer improvements in Europe and Latin America.	17 crisis countries across Asia, Europe, and Latin America	1990s
Sufian and Habibullah	2010	Loan-loss provisions, overhead cost	Panel data	Indonesian banks were more profitable before the Asian financial crisis; diversification and capitalization helped profitability, whereas size and high overheads hurt it.	Indonesia	1990 - 2005
Song and Lee	2012	Cash holdings, investment, payouts	Long-term panel analysis	East Asian firms increased cash holdings after the Asian financial crisis mainly by cutting investment, reflecting a stronger precautionary motive.	8 East Asian countries	1990 - 2006; 1997 - 1998

Empirical Evidence

This section introduces the dataset and provides a descriptive overview of ROA across the eight sectors before presenting the empirical analysis.

Data

The dataset was derived from GuruFocus[1] and utilized panel data covering eight different sectors listed on the Stock Exchange of Thailand (SET) for the time period of 2001-2024. The dependent variable is Return on Assets (ROA) which measures the profitability of the sector relative to total assets (Net Profit/Total Assets). The independent variables are two different dummy variables for the Global Financial Crisis and the COVID-19 pandemic. These dummy variables take 1 during the event period and take 0 before and after the event period. The event periods for the Global Financial Crisis and the COVID-19 pandemic are defined as 2008-2009 and 2020-2021, respectively. The dataset consists of 192 observations.

Descriptive Statistics and ROA Trends

Figure 1. ROA Trends by Industry (2001-2024)

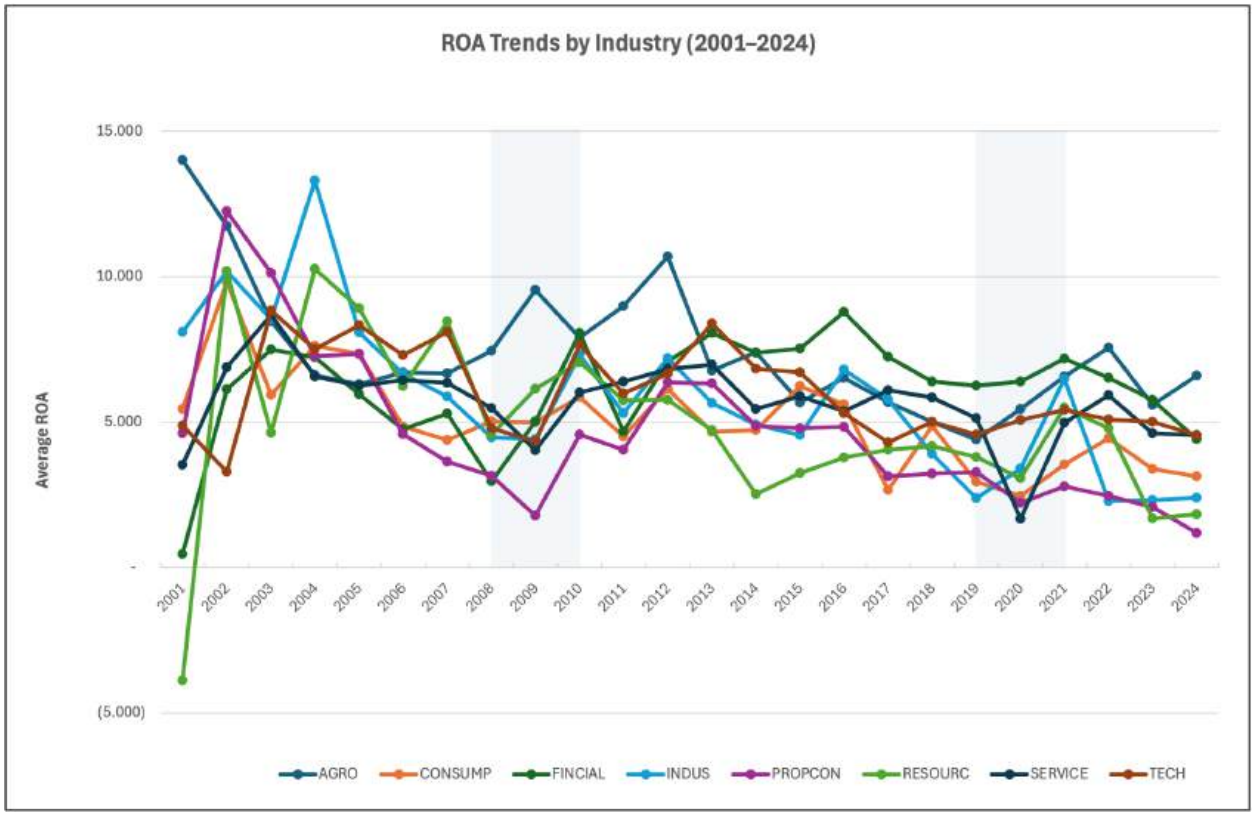
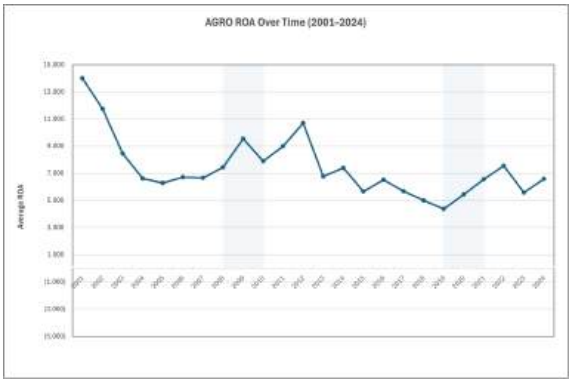
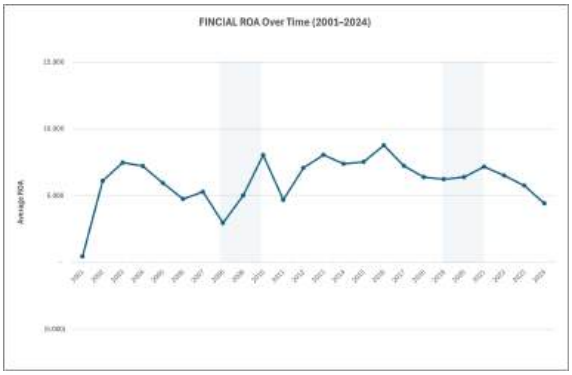
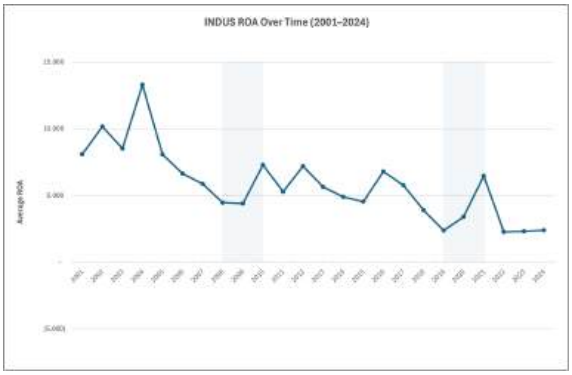
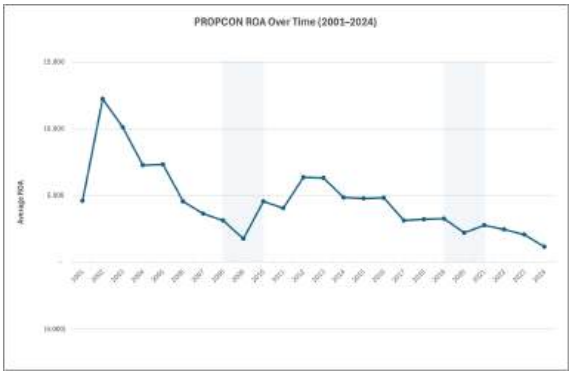
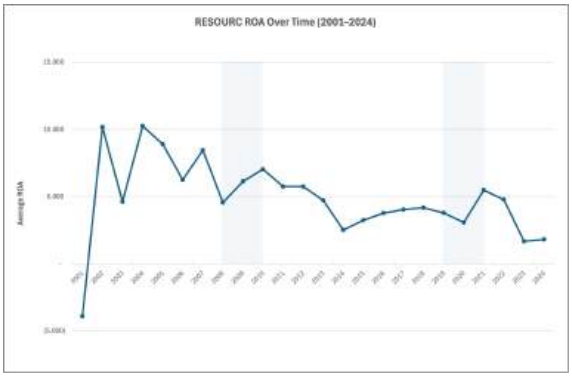


Figure 1 illustrates ROA trends for eight sectors listed on the Stock Exchange of Thailand (SET) over the period 2001-2024. The lines indicate that profitability was relatively high and volatile

[1] https://www.gurufocus.com/

during the early 2000s, declined around the Global Financial Crisis, and recovered to moderate levels in the years following. Around the COVID-19 period, most sectors experienced a drop in ROA, followed by a partial recovery during the early 2020s.

Figure 2. Sector-Specific ROA Trends (2001–2024)

Panel a. AGRO	**Panel b.** CONSUMP
	
Panel c. FINCIAL	**Panel d.** INDUS
	
Panel e. PROPCON	**Panel f.** RESOURC
	

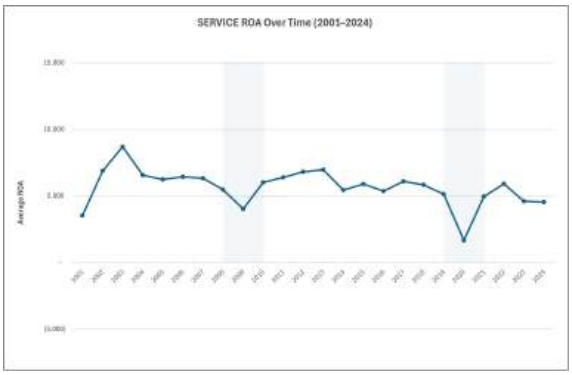
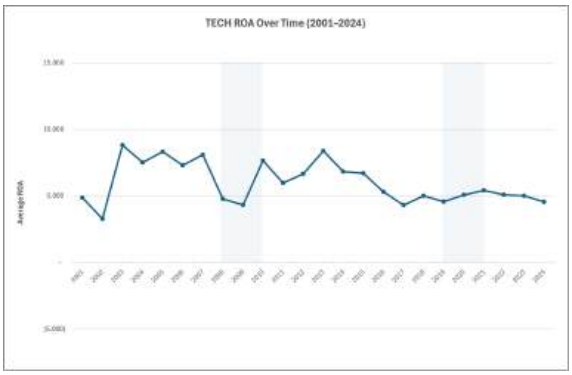
Panel g. SERVICE	**Panel h.** TECH
	

Figure 2 presents ROA trends for each sector individually from 2001 to 2024. Panels a–h show that Agriculture & Food (AGRO) and Resources (RESOURC) maintain relatively higher and more stable ROA over time, while Property & Construction (PROPCON) and Industrials (INDUS) had sharper declines around the GFC and weaker recovery afterwards. Consumer Products (CONSUMP), Financials (FINCIAL), Services (SERVICE), and Technology (TECH) exhibit moderate ROA with visible drops during the GFC or COVID-19 periods and mixed post-crisis recovery.

Empirical Model

$$ROA_t = \beta_0 + \beta_1 Dummy_{GFC} + \varepsilon_t \text{ (Model 1)}$$
$$ROA_t = \beta_0 + \beta_1 Dummy_{Covid19} + \varepsilon_t \text{ (Model 2)}$$

Where ROA_t is the dependent variable for observation "t" which covers the period of 2001 to 2024. The dataset covers eight different sectors, namely AGRO, CONSUMP, FINCIAL, INDUS, PROPCON, RESOURC, SERVICE, and TECH. The independent variables are two binary variables for $Dummy_{GFC}$ and $Dummy_{Covid19}$ which take 1 during the event period and 0 otherwise.

β_0 is the constant term, representing the expected value of the dependent variable when the independent variable is 0. β_1 is the coefficient for the independent variable ($Dummy_{GFC}$, $Dummy_{Covid19}$) which shows how much the dependent variable (ROA_t) changes when the independent variable ($Dummy_{GFC}$, $Dummy_{Covid19}$) changes by 1 unit. ε_t is the error term, which represents the difference between the actual value and the predicted value of the model.

Regression Results

Table 2. Regression Analysis

	Model 1	Model 2
VARIABLES	ROA	ROA
$Dummy_{GFC}$	-0.8925*	
	(0.4710)	
$Dummy_{Covid19}$		-1.2950***
		(0.4642)
Constant	5.7723***	5.8059***
	(0.1770)	(0.1760)
Observations	192	192
R-squared	0.0114	0.0241

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

Table 2 presents the results of two different regression models where the dependent variable is ROA and the independent variables are two different dummy variables, $Dummy_{GFC}$ and $Dummy_{Covid19}$.

The paper uses t-values to define the significance level of all estimated parameters. The paper indicates the statistical significance of the result at the 1 percent level (***), 5 percent level (**), and 10 percent level (*). The sign in front of the coefficient suggests whether the dummy variables have a statistically positive or negative impact on the ROA. "-" represents negative impact, while "+" represents positive impact.

In Model 1, the dummy variable ($Dummy_{GFC}$) and ROA are negatively associated with a coefficient of -0.893. The findings suggest that the ROA is 0.893 points lower when $Dummy_{GFC}$ = 1 (during the GFC period) than when $Dummy_{GFC}$ = 0 (during the non-GFC period). This result is statistically significant at the 10 percent level.

In Model 2, the dummy variable ($Dummy_{Covid19}$) and ROA are negatively associated with a coefficient of -1.295. The findings suggest that the ROA is 1.295 points lower when $Dummy_{Covid19}$ = 1 (during the COVID-19 pandemic period) than when $Dummy_{Covid19}$ = 0 (during the non-COVID-19 pandemic period). This result is statistically significant at the 1 percent level.

The comparative analysis of sector performance before, during, and after the GFC is presented in Figure 3, Table 3, and Table 4 below.

The comparative analysis of sector performance before, during, and after the COVID-19 pandemic is presented in Figure 4, Table 5, and Table 6 below.

While the regression analysis provides evidence of the overall impact of the GFC and the COVID-19 pandemic on profitability, the following analysis examines how these effects differed across sectors before, during, and after each crisis.

Sectoral Comparison During the GFC

Figure 3. Average ROA Before, During, and After the GFC

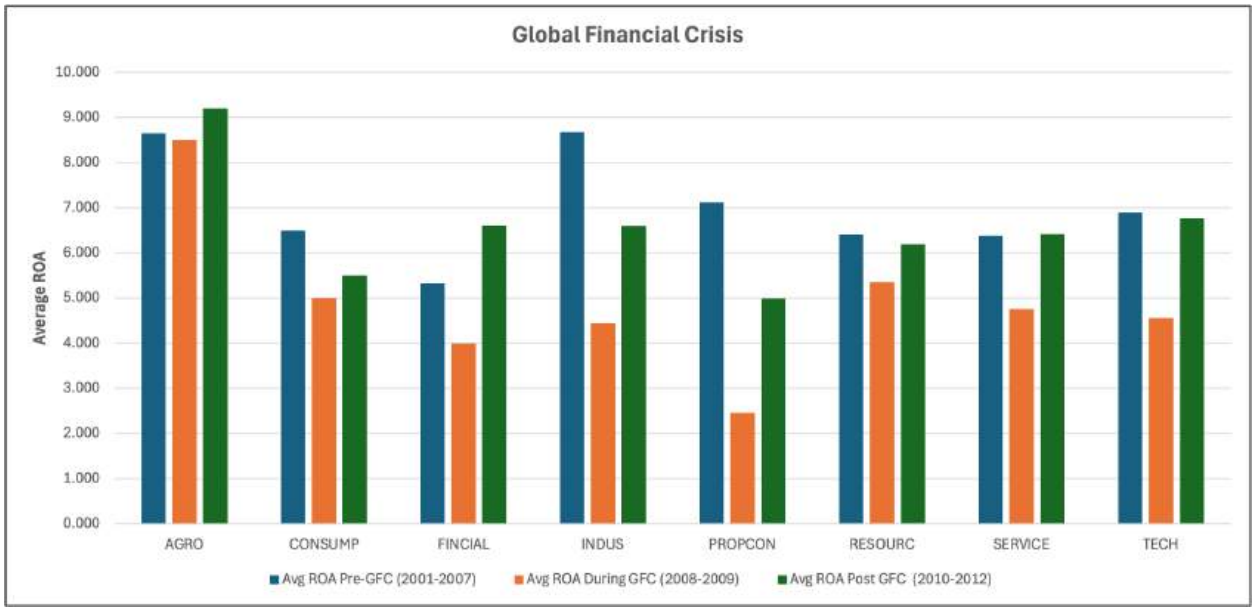


Figure 3 illustrates the comparative performance of eight different sectors before, during, and after the GFC. The evidence aligns with the regression findings, showing a decline in ROA during the GFC. It can be observed that sectors such as Agriculture were less affected by the GFC compared to other sectors, while the Property & Construction sector was the most affected by the GFC.

Table 3. Average ROA Before, During, and After the GFC

	Avg ROA Pre-GFC (2001-2007)	Avg ROA During GFC (2008-2009)	Avg ROA Post-GFC (2010-2012)
AGRO	8.646	8.496	9.198
CONSUMP	6.488	5.003	5.499
FINANCIAL	5.325	3.986	6.608
INDUS	8.675	4.438	6.603
PROPCON	7.116	2.457	4.996
RESOURCE	6.400	5.351	6.184
SERVICE	6.385	4.754	6.408
TECH	6.889	4.554	6.766

Table 3 highlights that not all sectors were equally vulnerable to the GFC. The largest decrease is in the Property and Construction sector (7.116 pre-GFC to 2.457 during) and the Industrial sector (from 8.675 pre-GFC to 4.438 during). In contrast, the Agriculture sector remained relatively stable (8.646 pre-GFC to 8.496 during). Post-GFC, most sectors recovered, with the Financial sector showing a strong increase to 6.608, exceeding its pre-GFC level.

Table 4. % Return on Assets Maintenance During and After GFC

	% of Pre-GFC ROA Maintained During GFC	% of Pre-GFC ROA Restored Post-GFC	During GFC Maintenance Rank	Post-GFC Recovery Rank
AGRO	98.27%	106.38%	1	2
CONSUMP	77.12%	84.76%	3	6
FINANCIAL	74.85%	124.09%	4	1
INDUS	51.16%	76.12%	7	7
PROPCON	34.53%	70.21%	8	8
RESOURCE	83.60%	96.62%	2	5
SERVICE	74.45%	100.35%	5	3
TECH	66.10%	98.21%	6	4

Table 4 further emphasizes the difference in resilience across the eight sectors. The data shows that Agriculture maintained most of its ROA during the crisis (98.27%), indicating strong resilience, while Property & Construction maintained the least (34.53%). In the post-GFC period, the Financial sector demonstrated the strongest recovery (124.09%), exceeding its pre-GFC level, whereas Property & Construction remained the weakest in recovery (70.21%).

Sectoral Comparison During COVID-19

Figure 4. Average ROA Before, During, and After the COVID-19 Pandemic

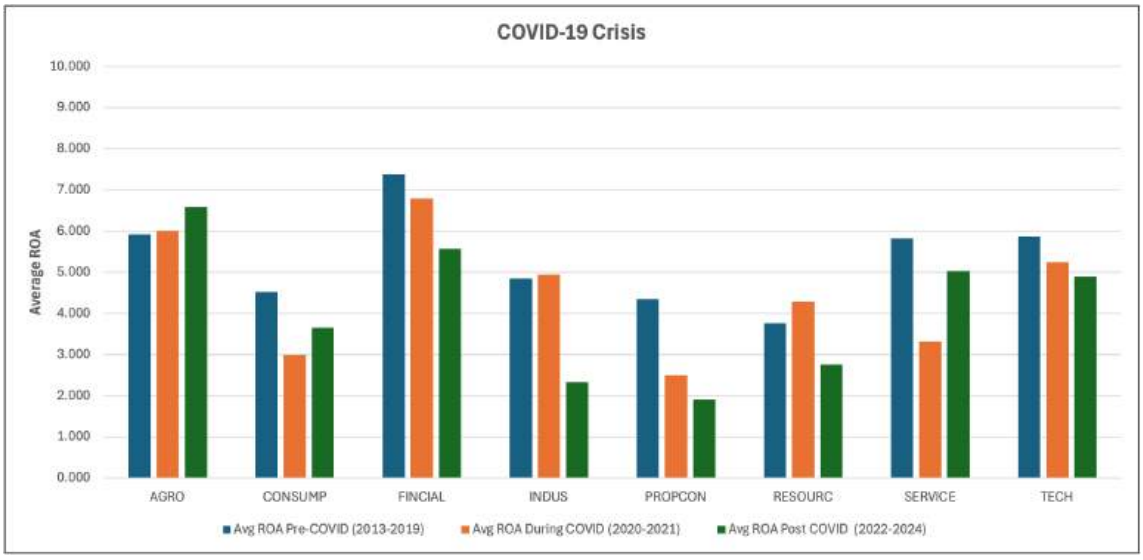


Figure 4 illustrates the comparative performance of eight different sectors before, during, and after the COVID-19 pandemic. The evidence aligns with the regression findings, showing an overall decline in ROA during the COVID-19 pandemic period. Agriculture was observed to be the most resilient among the sectors, as the average ROA post-COVID-19 was actually higher than before.

Table 5. Average ROA Before, During, and After the COVID-19 Pandemic

	Avg ROA Pre-COVID-19 (2013-2019)	Avg ROA During COVID-19 (2020-2021)	Avg ROA Post-COVID-19 (2022-2024)
AGRO	5.920	6.009	6.583
CONSUMP	4.526	2.997	3.649
FINANCIAL	7.379	6.789	5.570
INDUS	4.851	4.946	2.334
PROPCON	4.352	2.504	1.905
RESOURCE	3.755	4.281	2.762
SERVICE	5.823	3.319	5.024
TECH	5.874	5.242	4.884

Table 5 highlights substantial differences between how the sectors were affected by the COVID-19 pandemic. The largest decreases are in the Services sector (from 5.823 pre-COVID-19 to 3.319 during) and the Property & Construction sector (from 4.352 pre-COVID-19 to 2.504 during). In contrast, the Agriculture sector remained resilient, with ROA increasing slightly from 5.920 pre-COVID-19 to 6.009 during the pandemic, and the Resources sector also recorded a higher ROA during COVID-19 (from 3.755 to 4.281). Post-COVID-19, recovery was uneven: only the Agriculture sector exceeded its pre-COVID-19 level, rising to 6.583, while most other sectors declined further.

Table 6. % Return on Assets Maintenance During and After COVID-19 Pandemic

	% of Pre-COVID-19 ROA Maintained During COVID-19	% of Pre-COVID-19 ROA Restored Post-COVID-19	During COVID-19 Maintenance Rank	Post-COVID-19 Recovery Rank
AGRO	101.50%	111.20%	3	1
CONSUMP	66.21%	80.61%	6	4
FINANCIAL	92.01%	75.49%	4	5
INDUS	101.95%	48.11%	2	7
PROPCON	57.52%	43.76%	7	8
RESOURCE	113.99%	73.55%	1	6
SERVICE	56.99%	86.27%	8	2
TECH	89.25%	83.16%	5	3

Table 6 further emphasizes the difference in resilience across the eight sectors. The data shows that the Resources sector maintained the highest share of its pre-COVID-19 ROA during the pandemic (113.99%), indicating strong resilience, while the Services and Property & Construction sectors maintained the least (56.99% and 57.52%, respectively). In the post-COVID-19 period, the Agriculture & Food sector demonstrated the strongest recovery (111.20%), exceeding its pre-COVID-19 level, whereas Property & Construction remained the

weakest in recovery (43.76%).

Conclusion

The paper examines the effects of the Global Financial Crisis (GFC) and COVID-19 pandemic on the profitability of eight sectors listed in the Stock Exchange of Thailand (SET) from 2001 to 2024. This study shows that both crises significantly reduced corporate profitability. The results suggest that the COVID-19 pandemic had a larger impact on firms' profitability relative to the impact made by the GFC. Sectoral resilience and recovery varied greatly by sector. The Agriculture sector (AGRO) demonstrated higher resilience to crises, while the Property & Construction sector (PROPCON) experienced a larger amount of damage due to these crises. These findings help investors, managers, and policymakers construct more resilient businesses and practice better risk management in future crises. Future studies might include firm-specific information to help understand the impact of these major crises on individual firms. Overall, the findings suggest that different types of global crises do not affect all industries equally, highlighting the importance of sector-specific resilience strategies. Future studies could also use panel data approaches to take into account variation in both time and cross-section dimensions.

Appendix

Table 7. Data Definition

Indicator Name	**Definition**	**Source**
Return on Assets (ROA)	A profitability ratio calculated as (Net Profit/Total Assets) * 100	GuruFocus

Table 8. Sector Definition

Sector Code	**Definition**	**Source**
AGRO	Agriculture & Food Industry	Stock Exchange of Thailand (SET)
CONSUMP	Consumer Products	Stock Exchange of Thailand (SET)
FINCIAL	Financials	Stock Exchange of Thailand (SET)
INDUS	Industrials	Stock Exchange of Thailand (SET)
PROPCON	Property & Construction	Stock Exchange of Thailand (SET)
RESOURCE	Resources	Stock Exchange of Thailand (SET)
SERVICE	Services	Stock Exchange of Thailand (SET)
TECH	Technology	Stock Exchange of Thailand (SET)

Works Cited

Achim, M. V., Safta, I. L., Văidean, V. L., Mureșan, G. M., & Borlea, N. S. (2022). The impact of covid-19 on financial management: evidence from Romania. Economic Research-Ekonomska Istraživanja, 35(1), 1807–1832.

Alsharef, A., Banerjee, S., Uddin, S. J., Albert, A., & Jaselskis, E. (2021). Early impacts of the COVID-19 pandemic on the United States construction industry. International Journal of Environmental Research and Public Health, 18(4).

Bris, A., Koskinen, Y., & Pons, V. (2004). Corporate financial policies and performance around currency crises. The Journal of Business, 77(4), 749–796.

DevPrasad, Y. R. (2015). Return to profitability after a financial crisis. The International Journal of Business and Finance Research, 9(1).

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. Journal of International Financial Markets, Institutions and Money, 21(3), 307–327.

Farhat, A., & Hili, A. (2021). The performance of compliant stocks during the Covid-19 crisis. SSRN. https://ssrn.com/abstract=3916435

Horta, I. M., Camanho, A. S., & Da Costa, J. M. (2012). Performance assessment of construction companies: A study of factors promoting financial soundness and innovation in the industry. International Journal of Production Economics, 137(1), 84–93.

Kočišová, K. (2014, October). Profitability determinants and the impact of the global financial crisis. In 5th Central European Conference in Regional Science (pp. 396–406).

Lee, C. C., Chen, M. P., & Ning, S. L. (2017). Why did some firms perform better in the global financial crisis? Economic Research-Ekonomska Istraživanja, 30(1), 1339–1366.

Nazir, M. S., Safdar, R., & Akram, M. I. (2012). Impact of global financial crisis on banks' financial performance in Pakistan. American Journal of Scientific Research.

Olabisi, O. (2011). Causes and impacts of global financial crisis on the performance of Nigerian banks (a case study of selected banks). Journal of Business Economics and Management.

Peric, M., & Vitezic, V. (2016). Impact of global economic crisis on firm growth. Small Business Economics, 46(1), 1–12.

Phan, D. H. B., & Narayan, P. K. (2020). Country responses and the reaction of the stock market to COVID-19—A preliminary exposition. Emerging Markets Finance and Trade, 56(10), 2138–2150.

Sahi, A. (2019). Determinants of financial performance of insurance companies of USA and UK during the global financial crisis (2007–2016). International Journal of Accounting Research.

Shen, H., Fu, M., Pan, H., Yu, Z., & Chen, Y. (2020). The impact of the COVID-19 pandemic on firm performance. Emerging Markets Finance and Trade, 56(10), 2213–2230.

Song, K. R., & Lee, Y. (2012). Long-term effects of a financial crisis: Evidence from cash holdings of East Asian firms. Journal of Financial and Quantitative Analysis, 47(3), 617–641.

Sufian, F., & Habibullah, M. S. (2010). Assessing the impact of the financial crisis on bank performance: Empirical evidence from Indonesia. ASEAN Economic Bulletin, 245–262.

Thangaraj, R. K., & Chan, T. K. (2012). The effects of the global financial crisis on the Australian building construction supply chain. Australasian Journal of Construction Economics and Building, 12(3), 16–30.

Yap, B. C. F., Mohamed, Z., & Chong, K. R. (2014). The effects of the financial crisis on the financial performance of Malaysian companies. Asian Journal of Finance & Accounting, 6(1), 236–248.

Ying Lai, H., Rashid Abdul Aziz, A., & Khuan Chan, T. (2014). Effect of the global financial crisis on the financial performance of public listed construction companies in Malaysia. Journal of Financial Management of Property and Construction, 19(3), 246–263.

Zhang, R., Chen, L., & Sun, X. (2024). The impact of COVID-19 on the financial performance of the UK construction firms. Construction Management and Economics, 42(10), 867–885.

Understanding How Sneakers Act as a Social Currency Among Adolescents in India
By Dhairya Dalani

Abstract

Sneakers have evolved beyond functional athletic footwear to become symbols of identity, status, and social belonging. This study investigated whether sneakers function as a form of social currency among adolescents by examining first impressions, attention to visual cues, and perceived social influence of footwear. A survey incorporating vignette-based scenarios and the Sneaker Social Currency Scale was administered to 68 Indian adolescent participants. Descriptive analyses explored participants' first impressions and attention to footwear, while independent samples t-tests examined gender differences across six dimensions of sneaker social currency: status signalling, belonging and inclusion, peer surveillance, competence inference, social payoff and access, and perceived pressure. Both male and female participants consistently identified footwear as an important visual cue when evaluating others, although clothing, facial expressions, and overall appearance received greater attention in forming first impressions. Female respondents demonstrated slightly greater willingness to engage socially with the vignette character, whereas male respondents showed more neutral responses. The findings indicate that sneakers contribute to adolescents' social perceptions as symbolic markers of identity and belonging but operate alongside other appearance-related cues rather than independently. This study extends research on symbolic consumption by examining sneakers within adolescent social contexts and provides a foundation for future investigations into how consumer products shape identity, peer relationships, and social inclusion.

Keywords

Sneakers, Social currency, Adolescents, Footwear, and Peers

Introduction

Sneakers have evolved from functional athletic footwear into cultural symbols that communicate identity, status, and social belonging (Digitalcommons@uri and Dejesus). While originally designed to improve athletic performance, sneakers are now primarily consumed as lifestyle products, with their value increasingly shaped by branding, exclusivity, and cultural meaning rather than functionality (Gillath et al; Wright and Giles Sparrow). Over the past several decades, sneaker culture has expanded beyond sport to become a global phenomenon influenced by popular media, celebrity endorsements, music, and fashion, transforming sneakers into powerful symbols of personal identity and social status (Wright and Giles Sparrow).

The rapid commercialisation of the sneaker industry has further strengthened this transformation (Wright and Giles Sparrow; Berman). Major brands such as Nike and Adidas have adopted marketing strategies that emphasise limited releases, collaborations, celebrity endorsements, and artificial scarcity to increase product desirability and brand loyalty (Berman). These "drop culture" strategies encourage consumers to associate sneaker ownership with exclusivity,

prestige, and membership within particular social groups (Berman). Consequently, purchasing decisions are often driven by symbolic and emotional value rather than functional need, with consumers seeking products that communicate status, authenticity, and belonging (Berman; Liu). Within youth populations, sneakers have become particularly important social markers. Adolescence is characterised by heightened social comparison, identity development, and peer influence, making visible possessions especially meaningful. In sporting environments, where performance, appearance, and peer evaluation are highly salient, footwear may influence how athletes are perceived by teammates and competitors (Wright and Giles Sparrow). High-status sneakers may signal competence, commitment, and popularity, whereas older or less desirable footwear may contribute to negative social judgments or feelings of exclusion (Wright and Giles Sparrow). Thus, sneakers may function as a form of **social currency**, whereby ownership of particular footwear provides symbolic social value beyond its practical purpose (Berman). For instance, the $220 Hardaway Galaxy Foamposite is famous for its glow-in-the-dark look and high resale value (Berman). In 2012, crowds for a shoe release got so out of hand that over a hundred police officers, some on horseback, had to shut down a mall (Berman). Limited stock caused chaos, especially during yearly releases like the Air Jordan XI (Berman).

Although previous research has explored sneaker culture, brand loyalty, symbolic consumption, and consumer behaviour, existing studies have largely focused on fashion, marketing, or general consumer populations. Limited research has examined how sneakers function as social currency within youth sport environments or investigated whether footwear alone shapes perceptions of competence, social status, and belonging. Furthermore, few studies have developed empirical measures to assess the social significance of sneakers or experimentally examined footwear-based social bias.

Therefore, the present study aims to examine sneakers as social currency among young athletes by developing a Sneaker Social Currency Scale and using vignette-based scenarios to investigate whether footwear influences perceptions of competence, status, and social inclusion. By integrating perspectives from sports psychology, consumer behaviour, and youth identity, this study seeks to provide new insights into the social meanings attached to athletic footwear and their implications for peer relationships and consumption behaviour.

Methods

Tools Used

The current study used a quantitative approach with an image-based questionnaire. The questionnaire was in English and aimed to understand how individuals in India form social judgements and emotional responses based on the type of shoes worn by others. It consisted of 28 questions spread across 7 sections, with each section containing 3 image-based questions. Each image presented to participants depicted a person wearing a distinct style of footwear, and respondents were asked to evaluate the individual in the image across a range of perception-based indicators. Responses were collected using a 5-point Likert scale, ranging from

1 (“Strongly Disagree”) to 5 (“Strongly Agree”), which allowed for the quantification of participants’ attitudes and perceptions towards the individuals depicted. Tallying was additionally used to record the frequency of specific judgments made by respondents across the image-based questions, providing a clearer picture of patterns in social perception based on footwear type. The image-based format was selected to reduce response bias and to allow participants to form instinctive, visual judgements, thereby more accurately capturing implicit social attitudes.

Ethical Considerations

Before data collection, informed consent was clearly outlined at the start of the survey and obtained from all participants involved in the study. Completing the questionnaire indicated agreement to participate. Participants retained the right to withdraw from the survey at any point without consequence. All information provided would remain confidential and would not be disclosed to any third parties. Access to data files would be restricted solely to the researcher. Potential risks and benefits were communicated to respondents at the outset of the study. All responses would remain anonymous, and data would be de-identified when presented in the study. The researcher's contact details were also provided, should participants have any questions or concerns regarding the project.

Sample Characteristics

The survey had a total of 75 responses, out of which 7 responses were excluded as they did not fall in the right answering group due to a form error, making the final sample size 68. All participants were Indians, aged from under 10 to over 21 years. Among the respondents, 61.8% were male and 38.2% were female. The respondents were divided into two groups based on their gender, and they answered questions based on pictures of their own gender.

Data Collection Procedure

The primary data collection instrument was a structured online questionnaire developed using Google Forms. This platform was selected due to its accessibility, ease of use, and ability to automatically record and organise responses in real time. Google Forms ensured consistency across participant responses and simplified the process of quantitative analysis. The use of digital tools reduced manual errors, maintained data accuracy, and improved the overall reliability and efficiency of the research process. WhatsApp and email were used to distribute the form, and the survey was administered in English to ensure ease of access. The survey was distributed digitally via WhatsApp and email, and participation was entirely voluntary and anonymous.

Data Analysis Strategy

The data analysis for the present study focuses on understanding how sneakers act as a social currency among young adolescents. Descriptive statistics were first used to summarise the cleaned data, including demographic characteristics. Graphical representations like bar graphs

were used to present the data visually and highlight important key points. Further analysis was carried out with an independent sample t-test using the Social Science Statistics calculator. The t-test was conducted to examine differences between male and female respondents. A p-value of 0.05 was used as a threshold for statistical significance.

Results

The results of the following section are divided into three categories: sneakers as a social currency among female adolescents, sneakers as a social currency among male adolescents and differences between male and female respondents.

Sneakers as a social currency among female adolescents

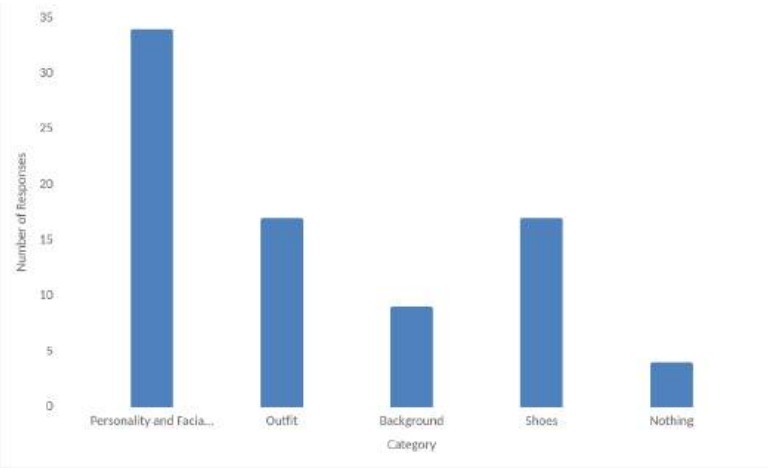


Figure 1: graphical representation of what was the female respondent's first impressions

In Figure 1, it can be observed that female respondents selected personality and facial expressions as their first impression (34), making it the highest response by a considerable margin. Outfit and shoes both received 17 responses each, while background had 9 responses, and nothing had 4 responses. This suggests that female respondents placed greater emphasis on facial and personality cues when forming a first impression, though footwear and outfit still contributed meaningfully.

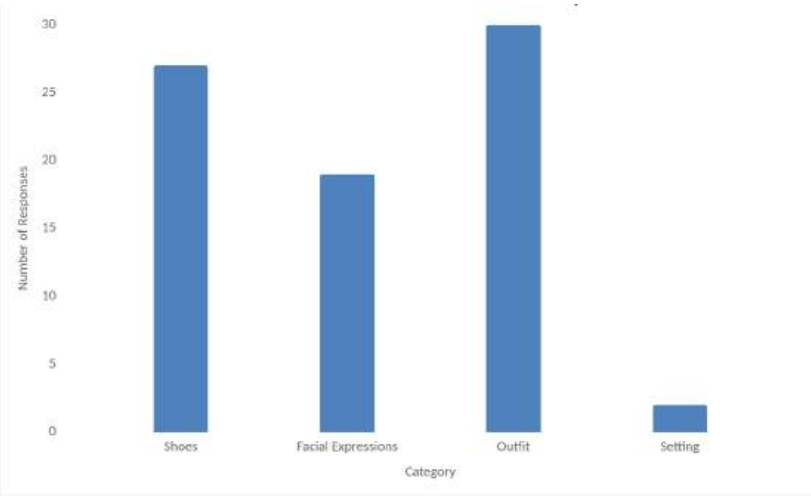


Figure 2: Graphical representation of what element of the pictures stood out to the female respondents

In Figure 2, outfits stood out the most to female respondents (30), followed by shoes (27) and facial expressions (19). The setting received very few responses (2). This indicates that female respondents were primarily drawn to the clothing and footwear of the person in the stimulus image, with the surrounding setting having very little impact on what they noticed first.

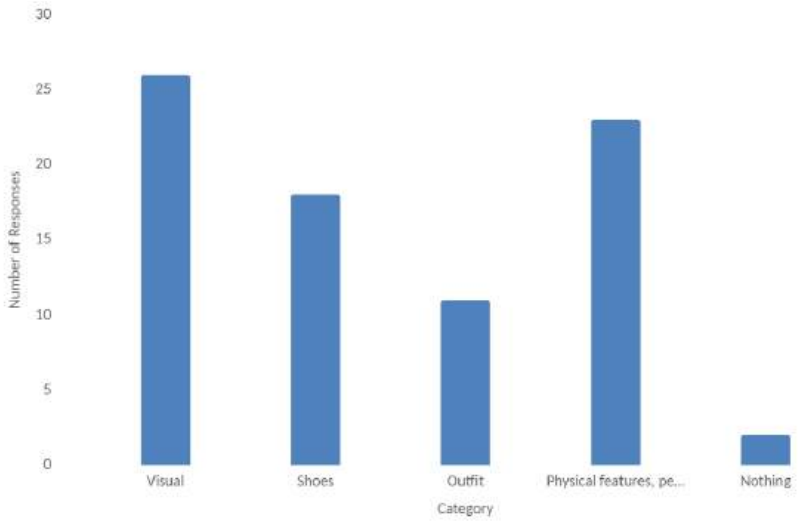


Figure 3: Graphical representation of what caught the female respondent's attention when viewing the picture

In Figure 3, it can be observed that visual aspects received the highest number of responses from female respondents (26), followed by physical features (23) and shoes (18). The outfit received approximately 11 responses, while nothing had the fewest (2). This suggests that female respondents were largely drawn to the overall visual presentation of the person, with shoes being a notable contributing factor to what caught their attention.

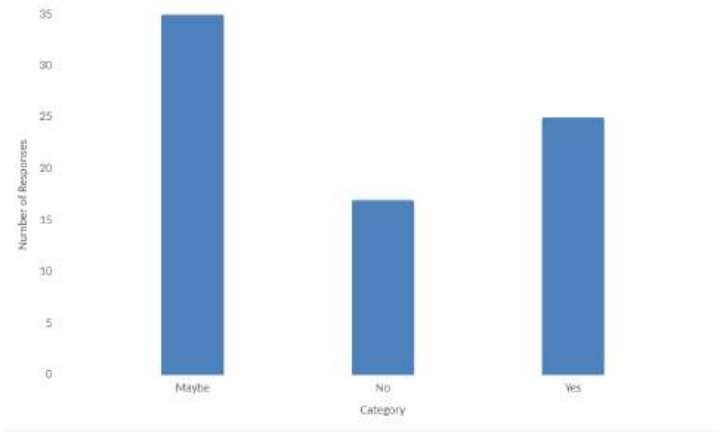


Figure 4: Graphical representation of the choice female respondents would make in knowing the person in the image

In Figure 4, it can be observed that Maybe received the highest number of responses among female respondents (34), followed by Yes (25) and No (approximately 16). Unlike male respondents, in whom no one outweighed the other, female respondents showed a slightly more positive inclination to want to know the person. This trend suggests that the appearance of the person in the stimulus image, including their footwear, had a relatively more positive influence on female respondents' social interest.

Sneakers as a social currency among male adolescents

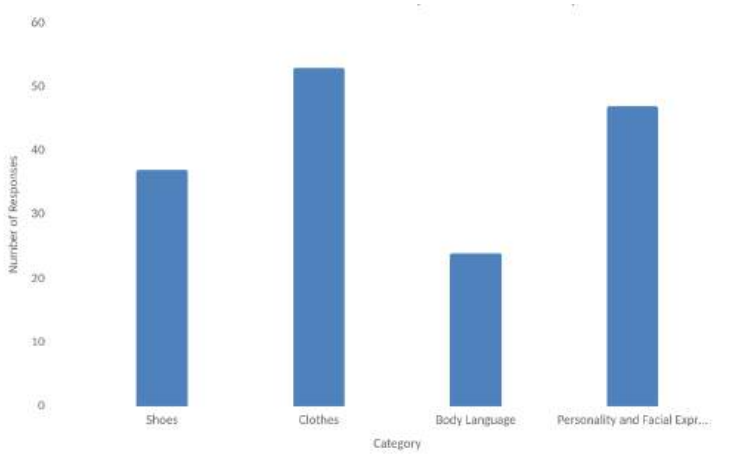


Figure 5: Graphical representation of what was the male respondent's first impressions

In Figure 5, it can be observed that among male respondents, clothes received the highest number of responses (52), followed closely by personality and facial expressions (47). Shoes also received a notable number of responses (37), while body language had the fewest responses (24). This suggests that male respondents primarily formed their first impressions based on

clothing, though footwear still played a significant role in shaping their initial perception of the person.

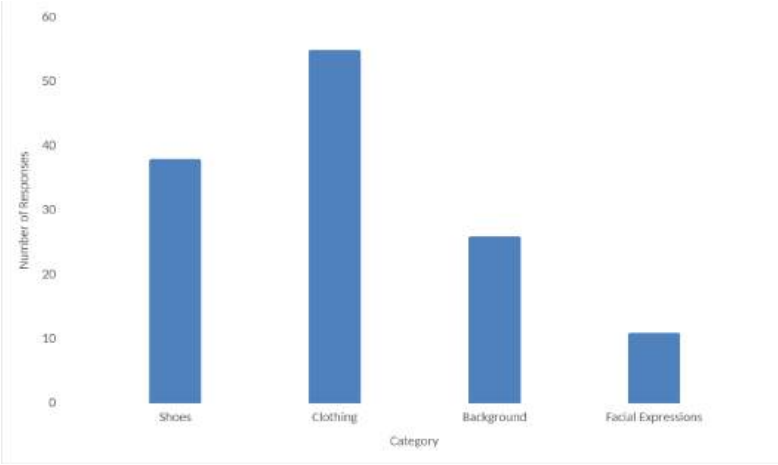


Figure 6: Graphical representation of what element of the pictures stood out to the male respondents

In Figure 6, clothing stood out the most to male respondents (55), followed by shoes (38) and background (26). Facial Expressions received the fewest responses (approximately 11). This indicates that among males, clothing and shoes were the dominant visual elements that immediately captured attention when viewing the stimulus images.

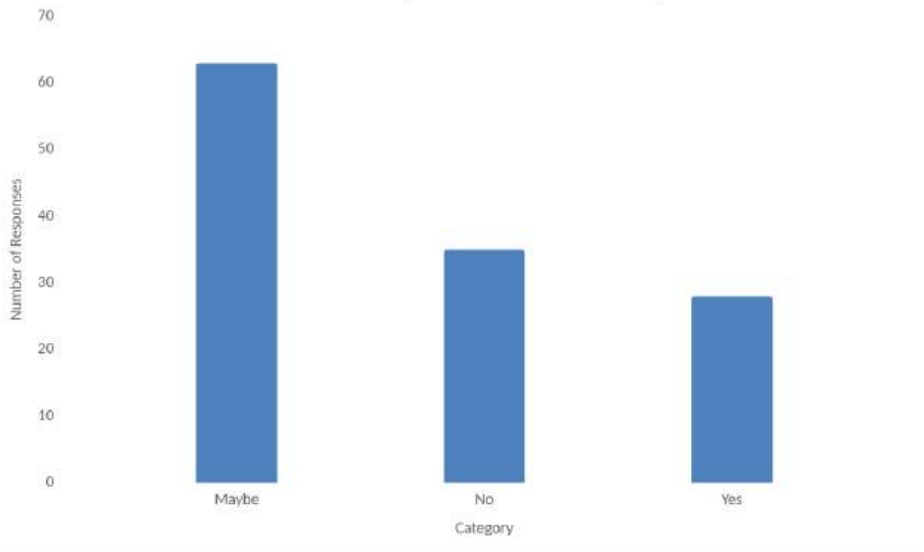


Figure 7: Graphical representation of the choice male respondents would make in knowing the person in the image

In Figure 7, it can be observed that the majority of male respondents selected Maybe (63), suggesting uncertainty rather than a clear positive or negative reaction. No received approximately 35 responses, while Yes had the fewest responses (28). This pattern indicates that the footwear and overall appearance of the person in the stimulus image did not strongly encourage male respondents to want to engage with the person further.

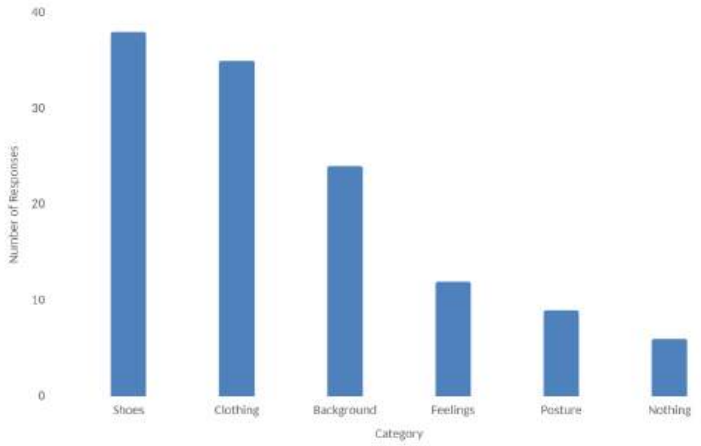


Figure 8: Graphical representation of what caught the male respondent's attention when viewing the picture

Figure 8 depicts that shoes received the highest number of responses (38), followed closely by clothing (35) and background (24). Feelings had approximately 12 responses, posture had approximately 9, and nothing had the fewest responses (6). This shows that for male respondents, the shoes and clothing of the person in the stimulus image were the primary reasons certain elements caught their attention, reinforcing the significant role footwear plays in visual perception.

Sneakers as a social currency: differences between female and male adolescents

Table 1: Independent t-test analysis of different sub scales by gender

Variable	Categories	N	Mean	Variance	t	*p*
Status Signalling	Female	26	13.65	12.72	0.45	0.65
	Male	26	14.12	14.43		
Belonging & Inclusion	Female	26	11.27	12.28	1.63	0.05
	Male	26	12.96	15.64		
Peer Surveillance	Female	26	13.04	12.28	-0.43	0.67
	Male	26	13.54	22.02		
Competence Inference	Female	26	11.42	10.33	0.17	0.35
	Male	26	10.42	19.13		
Social Payoff & Access	Female	26	11.54	11.78	-0.80	0.43
	Male	26	12.46	22.74		
Perceived Pressure & Behaviour	Female	26	9.81	9.68	-0.99	0.33
	Male	26	11.00	28.24		

*Note: $p<0.05$**

In Table 1, independent sample *t*-tests were conducted to examine gender differences across the six dimensions of the Sneaker Social Currency Scale. No statistically significant gender differences were observed for status signalling, $t(50) = 0.45$, $p = .65$, peer surveillance, $t(50) = -0.43$, $p = .67$, competence inference, $t(50) = 0.17$, $p = .35$, social payoff and access, $t(50) = -0.80$, $p = .43$, or perceived pressure and behaviour, $t(50) = -0.99$, $p = .33$. Male participants reported slightly higher mean scores than female participants for status signalling ($M = 14.12$ vs. 13.65), belonging and inclusion ($M = 12.96$ vs. 11.27), peer surveillance ($M = 13.54$ vs. 13.04), social payoff and access ($M = 12.46$ vs. 11.54), and perceived pressure and behaviour ($M = 11.00$ vs. 9.81), whereas female participants reported a marginally higher mean score for competence inference ($M = 11.42$ vs. 10.42). However, these differences were not statistically significant ($p > .05$).

Discussion

The present study explored whether sneakers function as a form of social currency among adolescents by influencing first impressions and social perceptions. Across both genders, participants consistently identified footwear as an important visual cue; however, clothing, facial expressions, and overall appearance were generally more influential in shaping first impressions. These findings suggest that sneakers contribute to impression formation but operate alongside other visible appearance cues rather than independently.

The findings support **symbolic consumption theory** (Gillath et al.), which proposes that consumer products communicate identity, status, and group membership beyond their functional purpose. Material possessions, particularly branded products, are often used to express social identity and gain social recognition. Although footwear alone did not determine participants' evaluations, its consistent role in attracting attention indicates that sneakers act as symbolic markers that contribute to adolescents' perceptions of others.

Interestingly, no significant gender differences were found across any dimension of the Sneaker Social Currency Scale. This suggests that the symbolic meaning attached to sneakers is similarly perceived by both male and female adolescents. The increasing influence of social media, celebrity culture, and global sneaker marketing may have contributed to the widespread adoption of sneakers as a shared cultural symbol rather than a gender-specific status marker. These findings have practical implications for schools, sporting organisations, and marketers. Since visible consumer products may shape peer perceptions and social belonging, educators and coaches should promote inclusive environments that emphasise individual ability rather than branded appearance. Similarly, marketers should recognise that campaigns promoting exclusivity and status may unintentionally reinforce social comparison and appearance-based pressure among young consumers.

This study has several limitations. The use of self-report measures and vignette-based scenarios may not fully reflect real-world behaviour, and the relatively small sample size may

have limited the ability to detect subtle gender differences. Furthermore, the study did not examine the influence of specific sneaker brands, price, or limited-edition releases. Future research should investigate these factors using larger, more diverse samples and experimental designs to better understand how sneaker culture shapes identity, belonging, and consumer behaviour among adolescents.

Conclusion

This study explored whether sneakers function as a form of social currency among adolescents by examining first impressions, social perceptions, and gender differences. The findings indicate that sneakers contribute to how adolescents perceive others, although clothing, facial expressions, and overall appearance also play important roles in shaping these judgments. While descriptive responses suggested that footwear is a noticeable component of first impressions, no statistically significant differences were observed between male and female participants on measures of sneaker social currency. These findings suggest that sneakers operate as symbolic markers of identity and belonging rather than independent determinants of social evaluation. As sneaker culture continues to expand through social media, celebrity influence, and consumer marketing, understanding the social meanings attached to footwear becomes increasingly important. Future research should continue to investigate how branded products influence adolescent identity, peer relationships, and social inclusion across diverse populations.

Works Cited

Berman, Aidan. Author Biography the Explosion of Sneaker Culture. 2023, digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1221&context=forum.

Digitalcommons@uri, Digitalcommons@uri, and Gianni Dejesus. A CULT CLASSIC: Sneaker Culture and its Influence on a Cult Classic: Sneaker Culture and its Influence on Consumer Behavior. digitalcommons.uri.edu/cgi/viewcontent.cgi?article=3285&context=theses.

Gillath, Omri, et al. "Shoes as a Source of First Impressions." Journal of Research in Personality, vol. 46, no. 4, Aug. 2012, pp. 423–30, https://doi.org/10.1016/j.jrp.2012.04.003.

Liu, Hongyu. "Development of the Emerging Shoe Market and Its Problems." Www.atlantis-Press.com, Atlantis Press, 15 Dec. 2021, pp. 1270–73, https://doi.org/10.2991/assehr.k.211209.205.

Wright, Michael, and Giles Sparrow. Marine Life : From Tropical Reef Fish to Mighty Sharks. Amber Books, 2016.

Global Stagflation and International Relations By Mikul Jain

Introduction

This literature review analyzes the connection between global stagflation and international relations through eight different scholarly works published between 1980 and 2022, examining stagflation's role in a country's foreign relations. Stagflation is a major worldwide issue that originated mainly in the 1970s due to the OPEC oil shock and stems from slow economic growth and high inflation. In macroeconomics, the concept of stagflation is generally very broad, and its causes and spread across the world are likely more than just supply shocks. Many scholars have often linked stagflation to global economic interdependence; however, others often argue that issues in domestic policy have a more significant role. Likewise, the literature in this paper remains divided over whether stagflation is an international phenomenon that emerges through economic interdependence or whether it is a domestic phenomenon that stems from a country's national policy. Taken together, this reveals an ongoing debate that serves as a central focus in this paper. The literature revolves around five main themes: research on the origins and spread of global stagflation, research on international institutions, research on the politics of stagflation, research on the US economy, and research on the European economy.

Research on the Origins and Spread of Global Stagflation

Many scholars and historians point to the OPEC crisis as the primary cause of global stagflation in the 1970s. In large part, this was the focus of many inflationary pressures not only in the United States but also worldwide. It is certainly true that the spikes in oil and natural gas prices were largely and directly a result of the petroleum-exporting world placing embargoes on oil exports to Western nations. However, research conducted by Robert Barsky and Lutz Kilian in "A Monetary Explanation of the Great Stagflation of the 1970s" argues that the OPEC crisis simply isn't enough to explain the origins of stagflation in the 70s. Instead, their research largely looks at worldwide monetary policies that further pushed inflationary pressures in global economies and which fueled stagflation. While the oil crisis is certainly a major cause of stagflation, it is necessary to understand the larger underlying issues in the 70s that truly made it a worldwide issue.

The research that Barsky and Kilian contend from a monetary policy perspective indicates that several central banks around the world pursued an expansion in their monetary policies—but at a time of already rising consumer prices. Keynesian macroeconomic principles generally tend to avoid expansionary policy during a time of inflation. By expanding their monetary base, economies in the 70s saw massive money supplies that simply raised consumer prices everywhere, leading to already higher-than-usual inflation and hence fueling global stagflation.

Their research also emphasizes the perspective that stagflation in the 70s didn't just come through supply shocks as many would assume, instead looking towards stagflation as a means for individual countries. Several large, advanced economies, like the United States as well as

Western European nations at the time, had very loose monetary policies, meaning that they excessively relied upon low interest rates and a large money supply to promote economic growth. In the long run, this only elevated consumer prices, feeding inflation amidst the crisis.

When it comes to commodity prices, usually most goods increase in price over an extended period of time. However, as seen, large-scale expansionary policy led to more severe consumer price increases. Supply shocks mainly cause prices of resources like oil and gas to increase directly, not commodity goods. The idea here is that external policy-related issues naturally led to inflation, causing prices for nearly every good to increase, which decreased people's purchasing power for those goods. In return, it fostered slow and sluggish economic growth in the mid-70s.

Additionally, Barsky and Kilian's research also looks at inflationary expectations during the stagflation era. Businesses in the 70s anticipated inflation increases, so in return, these businesses increased their costs as well as their wages. Naturally, this led to a long-term increase in the economic price level, triggering high amounts of inflation.

Another major event that led to global stagflation, as Barsky and Kilian observe, was the fall of the Bretton Woods system in the 1970s, in which fixed exchange rates were replaced with flexible rates. This issue largely dealt with the idea that it would allow for national currencies to fluctuate in value based on their demand and interest rates. This way, central banks were inclined to monetary autonomy and fight growing unemployment levels in the 70s. However, overexpansion via lowered interest rates only continued to fuel spending and led to increased inflation within economies.

While individual countries' monetary policies led to inflation, global stagflation was truly spread worldwide due to the fact that global economic monetary policies were largely connected to one another. Before the fall of the Bretton Woods system, central banks around the world were often dollar-dependent and largely followed one synchronized business cycle. Due to several economies pursuing similar monetary policies and becoming more integrated in the globalized economy, stagflation-like pressure sprang up across multiple countries. This helps to understand that international economic relations were heavily constrained by the global business cycle asserted by the US dollar.

From the origins and spread of stagflation in the 1970s, it is now understood and recognized that stagflation is certainly hard to counter, and its effects are widespread everywhere around the world. The research that Barsky and Kilian provide suggests that interconnected monetary conditions may have led to the rise of stagflation across multiple advanced nations. Barsky and Kilian's findings weren't quite understood very well in the latter half of the 20th century; however, they are quite relevant today. Battling stagflation requires strong anti-inflation monetary policy, which will truly build trust within an economy. However, while this literature emphasizes the importance of global monetary conditions, it leaves unresolved the question of whether global stagflation spreads through international economic interconnectivity or domestic policy choices in individual countries.

Research on International Institutions

International Institutions also play a big role in the way inflationary pressures and economic slowdown spread across nations. International Institutions are generally very broadly defined as a large organization or bodies that handles international affairs. Several well-known institutions include the UN (United Nations) as well as the WTO (World Trade Organization). Research, as seen by John Duffield in "What are International Institutions?," highlights the idea that these international institutions' main role in the world is to create predictable expectations around the world. They try to make sure that all nations are economically included and establish trade fairly by following rules and norms. When such playing fields are "predictable", the idea is that nations enjoy the benefits of trade and positive relations with one another. Their long-term goal is to establish stability.

Duffield also goes on to say that these institutions try to reduce "uncertainty". Uncertainty revolves around several global issues, which can be political, social, or economic. Stagflation, for example, would be an example of a global "uncertainty", which is seen as a worldwide crisis. International institutions often make it their duty to step up and ensure that countries are more willing to cooperate and follow norms and rules that help reduce uncertainty. And, as seen in historical examples, the role of these institutions in the past has actually been taken adequately. For example, in 1974, the United Nations established the Charter of Economic Rights and Duties of States, which established a guiding framework for sovereign equality for developing nations to better regulate foreign investment and MNCs amidst global stagflation. This understanding helps establish the idea that the role of institutions attempts to moderate relations amongst nations during economic crises.

In other words, what international institutions really aim for is the idea of "anarchy". Although no central worldwide government can act as an entity and govern every sovereign nation, these institutions attempt to do so and rely upon an anarchist framework to ensure stability. Duffield explains how international institutions ensure open communication and cooperation. While it is often the idea that institutions "attempt" to do so, it can come with the problem that these institutions often do too little, with nations showing little cooperation and interest, or do too much and burden themselves trying to find a medium amongst nations.

Furthermore, rules and norms of international institutions are also a key component of establishing anarchy. These rules and norms are not "policies" but rather more like guidelines. For example, the WTO (World Trade Organization) ensures that there are no underpricing or overpricing issues when countries facilitate trade. As Duffield explains, these norms try to determine appropriate behaviors for states. Existing institutional literature exclaims that in the event of global recession or inflation, these institutions encourage fair pricing and regulatory norms so that nations around the world can trade and interact with one another without having to show concern for an individual nation's behavior. Likewise, the WTO may ensure adequate markets in the event of a supply shock. Since nations always try to pursue economic self-interest, institutionalist scholars argue that the WTO can help oversee the behavior of states through setting rules and norms.

Institutions like the WTO also attempt to embrace globalization and open trade across the world. The research presented by Kalim Siddiqui in "International Trade, WTO and Economic Development" highlights that the WTO specifically emphasizes modern-day globalization across the world. However, Siddiqui also maintains that globalization can unintentionally make less developed nations (LDCs) vulnerable. While it is certainly true that the benefits of globalization, like development and economic growth, for example, are prevalent in many cases, it is unevenly distributed amongst the less developed nations. What this could trigger is a supply shock if resources like natural gas and oil, as well as other forms of energy, are not readily available and bring slow economic growth in less developed nations, which in the long run could only feed stagflation if it were to occur. While Siddiqui's research does not directly look at stagflation, such vulnerabilities may be relevant in stagflationary pressures.

While Duffield sees the WTO as a more positive and stability-seeking body, Siddiqui highlights some of the WTO's weaknesses and how they disrupt international order and economic stability amongst the nations of the world. When considering that the WTO's role is to promote open-market reforms, it doesn't exactly account for the negatives that these LDCs face at the hands of globalization. Issues that are often mentioned are the fact that wealthier, more technologically advanced nations often try to pursue their own self-interest, commonly influencing domestic employment levels in developing nations, which are often also pressured into globalizing. This fosters high unemployment levels and economic slowdown in several LDCs.

Despite the weaknesses of the WTO highlighted by Siddiqui, it is often more persistent that international institutional bodies do try to ease the effects of global crises and bring stability. The findings from Duffield re-emphasize that institutions provide for long-term stability and establish necessary rules and norms to keep order. Numerous recent policies and guidelines by institutions like the UN and WTO have made this possible, especially in the post COVID-19 era. The role that institutions play is to extensively attempt to ease economic slowdowns (e.g., stagflation) in order to balance and provide a medium for global economic affairs.

Research on the Politics of Stagflation

It is quite important to mention that economics and politics are universally tied together, so not mentioning the political aspects of economic slowdowns does not capture what exactly occurs when nations across the globe interact with one another. That is where the political aspects of stagflation try to fill the gap between international economics and the political circumstances around the world. The research found in R. J. Barry Jones in "International Political Economy: Problems and Issues: Part I" explains how international relations can no longer just be understood through the means of diplomacy, but rather need to take economic realities into account as well. This field of study is known as international political economy (IPE).

A major argument that Jones incorporates is the idea that worldwide economies are interdependent. He reasons that interdependent economies allow for shocks originating in one

country to influence economic conditions elsewhere through trade and financial connections. Establishing this framework helps understand how and why stagflation in the 70s spread across the world. Since economies rely on trade and resources, the spread of economic impacts is persistent.

The research also tends to focus on the political role of multinational corporations (MNCs) as well. MNCs are large corporations that are usually based out of advanced, industrialized countries. These corporations are often responsible for FDI (foreign direct investment) in developing nations. Many of these corporations include Walmart and McDonald's, which employ several people across the world and often hold huge economic power within developing nations by directly integrating their workforce and setting up infrastructure. The research on the political and economic power of MNCs is highlighted by the works of Joseph LaPalombara and Stephan Blank in "Multinational Corporations and Developing Countries." They propose that MNCs often provide economic growth in LDCs by providing their populations with jobs as well as establishing infrastructure. However, they also argue that MNCs can pose some human rights concerns, and their role can largely shape international political decisions.

At the same time, Jones looks at how strong individual economies directly dictate global politics, too. In IPE, this issue is known as "high politics". Both the role of MNCs and high politics explain how worldwide multinational politics impact the economy of the world itself. Following stagflation in the 1970s, politicians and economists often look at these aspects to explain economic stress and how they impact cross-border interactions.

Together, the research of Jones as well as LaPalombara and Blank elaborates upon 3 major ideas: global north-south relations, dependence on foreign investment, as well as the inequality faced between advanced industrial countries (AICs) and LDCs. For example, Jones emphasizes that global north-south relations are often strained by the emphasis on high politics, which favors the interests of the global north. This, in turn, as LaPalombara and Blank would argue, can cause several MNCs to create economic dependencies that may place vulnerability on several LDCs during economic disruptions. This framework leads scholars and experts to often confidently say that inequalities in the global economy may leave several LDCs vulnerable during times of economic disruption, like stagflation, for example.

The work by Jones additionally notes examples of historical instances of how stagflation in the 70s changed the way governments pursue foreign policy. It explains that the oil crisis of 1973 led to several Western Nations in disagreement with each other after the collapse of the Bretton Woods system. These nations had self-interests and changed their foreign policy in order for their foreign affairs to account for trade. Many of these governments became more protectionist in order to protect their domestic economies from the spread of stagflation. Following the oil crisis, several governments, such as France and Italy, put in place protectionist measures-such as import restrictions and trade barriers against imports from countries like Japan-in order to protect their domestic industries. It called for a stronger and tighter negotiation process with the International Monetary Fund (IMF).

What Jones concludes is the idea that no individual government can truly look over its own economy, and that is where the need for further study in IPE must come in. When considering stagflation, for example, scholars know that it revolves around the way nations interact economically. However, what remains a deeper question is the idea of sovereignty. As mentioned, since no country can truly control its own economy, it must rely upon the policies of other states as well as the norms of international institutions. Some scholars would argue that economic interdependence can strain autonomy, which raises questions about the practical limits of economic sovereignty. While institutions like the UN attempt to uphold the sovereignty of nations, it remains unclear whether that sovereignty can be meaningfully preserved when nations face economic challenges such as stagflation and other foreign pressures. This remains an issue that the research on the political aspects of the global economy needs further investigation.

Research on the US Economy

Understanding economic issues and stagflation and their connection, the IR must also address the research that investigates the historical and present US economy and how it establishes itself in the vast globalized world economy. While many scholars do point to the oil crisis as the primary and leading cause of the 1970s stagflation in the United States economy, the research studied by Thomas D. Willett in "The United States and World Stagflation: The Export and Import of Inflationary Pressures" puts a different perspective on how stagflation really hit the American economy. Much of Willett's research emphasizes domestic policy issues as the cause of stagflation in the US. In the past, several US policymakers did ultimately blame foreign pressures for high amounts of inflation and slow economic growth in the 70s, but in a way, it was also the policymakers who impacted the start and spread of inflation in the domestic economy. However, unlike Barsky and Kilian's emphasis on international monetary conditions, Willett argues that domestic policies determined how severe stagflation was within individual nations.

One perspective that Willett offers that differs from other scholars is the idea that the global economy became highly interdependent. While that is certainly the case, it is often overemphasized. What he claims is the idea that American unemployment was overemphasized and exaggerated in the 1970s. The idea that large levels of unemployment in the United States were created simply from the oil supply shock does not justify it as the only reason for the stagflation seen in the 1970s. As previously mentioned, foreign pressures definitely do transfer over, but how much of an impact they actually have on a large economy is hard to answer. This is often a field where several scholars cannot directly give empirical evidence, as it requires extensive macroeconomic investigation, which can be quite difficult to conduct and investigate.

The same research also takes into account the idea that inflationary pressures in the US economy during the 1970s weren't necessarily exported to other parts of the world. Oftentimes, it is noted that countries blame the US for tremendous increases in global inflation levels. It is often quite the opposite. This perspective maintains that countries across the world experiencing high levels of inflation were likely due to their domestic policy issues. Even in terms of macroeconomics, inflation altogether cannot be "exported" from one country to another.

Conversely, Willett also argues that inflation was not directly imported either. Exchange rates and oil prices often misinterpret the problem. Instead, the view re-emphasizes domestic US policy issues. The policy at the time made it easy for the spread of foreign shocks. Many of the domestic policies in the United States at the time fueled economic expansion, but at a time when prices were already climbing. For example, the Nixon administration implemented a 90-day freeze on all wages in 1971. This created lower purchasing power for several Americans because prices were climbing, while their income was not changing. Additionally, the Federal Reserve mistakenly believed in lowering unemployment rather than fighting growing inflation, which severely backfired. So, when it comes to stagflation, domestic policy plays a big role in its impacts. It is nearly impossible to confidently say that the crisis was entirely due to domestic issues or foreign shocks, but at the same time, the exact extent to which both occurred is also hard to determine.

Another idea that scholars who research stagflation and international relations in terms of the US economy also look at is that exchange rates don't necessarily indicate economic health. Before the Bretton Woods system, the general belief was that the stronger a currency was, the less prone it was to economic failure and indicated strong economic health. However, after the Bretton Woods system of fixed exchange rates collapsed, the world needed a better way of explaining economic health. Since expansionary monetary policy in the 70s only inflated the currency, the dollar largely depreciated. Modern-day exchange rates cannot fully assess the health and stability of economies as they fluctuate based on demand, as well as domestic interest rates. Research done by Georgios Karras in "Exchange-Rate Regimes and the Effectiveness of Fiscal Policy" examines exchange rates in the United States and concludes that further empirical evidence is needed before exchange rates can be considered a reliable indicator of a country's economic health.

Willett also illustrates that the oil shock itself had a limited role in the stagflation within the US. The research highlights that other nations were only lightly impacted and did not see stagflation as severe as the US. What it comes down to is an individual country's economic structure. Foreign, international pressures in the United States were by all means an issue, but the same cannot specifically be said for others regarding US-based pressure. The research summarizes that the level of shocks comes from economic structures, not from the outcomes as a whole.

Another aspect of the research on the US economy looks at the politics behind the 1970s' stagflation crisis. Research by Alberto Alesina and colleagues in "A Model of the Political Economy of the United States" finds that the economy and elections are directly connected to each other. No matter how severe the economic crisis within a country is, voters always respond based on the economy. The 70s showed that the voter base wanted an urgent and quick remedy to the stagflation in the economy. Naturally, voters gave their votes to those who promised improvement in the economy. However, politicians and policymakers at the time needed to understand the economy at the time better to actually make accurate decisions. Alesina and colleagues suggest that policymakers often struggle to respond effectively to new and previously

unknown economic conditions, a challenge that was prominent during the stagflation era. When it comes to fighting global stagflation, countries cannot always rely upon the political decisions of others but rather need to carry out domestic policy effectively.

Research on the European Economy

While the research on the US economy gives a glimpse of what stagflation and IR looked like in the 70s, the research focused on the European economy looks more into some modern questions regarding stagflation and the global economy. The research presented by Markus Demary and Michael Hüther in "How Large Is the Risk of Stagflation in the Eurozone?" identifies some of the current risks of stagflation in the European economy. Already in the European economy, there are elevated levels of inflation. This largely stems from production costs for many European companies whose primary customers are domestic consumers. Additionally, there is also slow economic growth and recovery in Europe coming from a multitude of factors.

One particular aspect of the European Economy that Demary and Hüther note is the supply chain bottleneck that several European economies are facing. Many European tech firms rely on trade with several Asian markets for the supply of semiconductors. Recently, the European Union (EU) has made sure to restrict several imports coming from Asian ports, especially China, which is a major provider of semiconductors on the global market. The idea behind it is that these European nations do not want to be vulnerable to trade and logistics imbalances with China. The relation between firms' needs for semiconductors and political tension between Europe and China directly clashes, as limitations on semiconductor supply halt economic capital growth in Europe. This has also led to high costs for the automotive and electronics industries in Europe.

Furthermore, the research also observes modern energy shocks in Europe. The COVID-19 pandemic saw reduced energy usage and consumption in Europe due to lockdown and quarantine policies. Later, several European nations placed sanctions on major energy and natural gas exporters like Russia. The problem that really elevated the situation was that there was a low amount of energy in storage, which seriously made oil and natural gas scarce in the European market. As a result, it led to very large spikes in oil and natural gas prices recently in the European market. This has most definitely led to higher rates of inflation in Europe than anticipated.

Additionally, Demary and Hüther also incorporate the fact that Europe's higher carbon prices are driving up inflation. Premiums that are paid on greenhouse emissions have only proven to raise costs for several businesses and firms, which is leading to higher price levels for goods. It is also causing less capital growth in Europe, which is slowing economic progress. With both these issues occurring simultaneously, it seems like Europe could already be experiencing stagflation or is on the verge of falling into it.

The same research goes on to highlight Europe's challenging monetary policy and its exchange rates. With stagflationary symptoms already starting to appear, European monetary

policy is experiencing trouble. With inflation already high, attempts to increase interest rates will likely lower investment spending, leading to further economic slowdown. On the other hand, decreasing interest rates would only feed inflation further.

This leads to a labour market risk in Europe. The research by Demary and Hüther (2022) highlights that Europe might fall into a wage-price spiral, a condition where inflation within a country causes workers to demand higher wages. Once wages are adjusted, business costs rise, leading to more inflation. This creates a cycle in an economy where unstable inflation pulls prices higher and higher within a country. These conditions, as Demary and Hüther (2022) identify, may increase the risk of stagflation in Europe.

Lastly, the research by Demary and Hüther also dives into the "recession" or "stagnation" being seen in the European economy. It is already clear that Europe is experiencing rising inflation, but at the same time, demographics within the continent are also shaping its economy. The aging population in Europe is leading to fewer available workers to expand capital in Europe. In the long run, this will only cause Europe to have a smaller economic output in the global economy. The authors note that such persistent inflationary and demographic pressures may weaken Europe's economic performance, potentially affecting exchange rates and the competitiveness of trade. All of the following have the possibility of causing Europe to enter stagflation. The research cannot, however, answer when this will happen; likewise, with other aspects of research already mentioned, it is also hard to say whether this stagflation would spread internationally.

Research Gaps and Future Implications

There is no doubt that the existing literature provides key insight into the relation between IR and global stagflation. However, there are a couple of gaps in the research presented. Understanding these gaps can bring a thorough conclusion and summary on the connection between the two variables.

The most significant of these research gaps is the idea that the literature remains divided on the extent of stagflation as an international or domestic phenomenon. The research that Barsky & Kilian, Jones, and Duffield present portrays stagflation as more of an international phenomenon through the role of economic interdependence, worldwide monetary conditions, as well as institutional frameworks. On the other hand, researchers such as Willett placed a greater emphasis on domestic policy decisions in the 70s. Due to this, there is no clear scholar-based view on which factors actually play the most dominant role in the spread of stagflation.

Additionally, much of the research tends to focus on either stagflation in the 70s or modern, post-pandemic stagflationary issues like those seen in Europe. There is relatively little research that emphasizes and compares modern stagflationary risks and historical stagflation. Future research can dive into looking at modern examples of inflation and contractionary issues and compare them to the historical stagflation of the 70s.

Finally, there is very little empirical evidence that can actually say whether international or domestic issues had a greater extent on the spread of stagflation. Future research that can

accurately say which factor plays a more dominant role can provide greater insight into this relevant topic. As Willett demonstrates, this will take a very large-scale macroeconomic investigation to truly provide a definite answer; however, it remains a topic that future researchers and scholars can look into.

Conclusion

The studies and research mentioned provide a solid bridge between global stagflation and international relations. The literature often suggests that economic interdependence can allow for economic crises internationally; however, scholars remain divided over whether stagflation is spread through international economic relations or failures in a nation's domestic policy. Simultaneously, international institutions can also shape a reality where some nations see stagflation worse than others. At the same time, it is not always necessary that foreign influences like supply shocks are the cause of stagflation in a country. Oftentimes, it is domestic policy that can fuel an economic crisis within a country. The literature presented shows that global economic interdependence shapes the way stagflation arises. However, the scholars remain divided on the idea of whether stagflation is transmitted through international mechanisms or through decisions in domestic policy. Resolving this debate is one of the most important gaps in the literature. There really is no one clear explanation, and it will take a very deep, thorough, large-scale investigation of macroeconomic and worldwide political factors to firmly say which one is actually the case

Works Cited

Alesina, Alberto, et al. "A Model of the Political Economy of the United States." The American Political Science Review, vol. 87, no. 1, 1993, pp. 12–33. JSTOR, https://doi.org/10.2307/2938953.

Barsky, Robert, and Lutz Kilian. "A Monetary Explanation of the Great Stagflation of the 1970s." SSRN Scholarly Paper no. 214990, Social Science Research Network, 1 Feb. 2000. papers.ssrn.com, https://papers.ssrn.com/abstract=214990.

Demary, Markus, and Michael Hüther. "How Large Is the Risk of Stagflation in the Eurozone?" Intereconomics, vol. 57, no. 1, Jan. 2022, pp. 34–39. Springer Link, https://doi.org/10.1007/s10272-022-1025-x.

Duffield, John. "What Are International Institutions?" International Studies Review, vol. 9, no. 1, 2007, pp. 1–22. JSTOR, https://www.jstor.org/stable/4621775.

Jones, R. J. Barry. "International Political Economy: Problems and Issues: Part I." Review of International Studies, vol. 7, no. 4, 1981, pp. 245–60. JSTOR, https://www.jstor.org/stable/20096927.

Karras, Georgios. "Exchange-Rate Regimes and the Effectiveness of Fiscal Policy." Journal of Economic Integration, vol. 26, no. 1, 2011, pp. 29–44. JSTOR, https://www.jstor.org/stable/23000907.

LaPalombara, Joseph, and Stephen Blank. "Multinational Corporations and Developing Countries." Journal of International Affairs, vol. 34, no. 1, 1980, pp. 119–36. JSTOR, https://www.jstor.org/stable/24356344.

Siddiqui, Kalim. "International Trade, WTO and Economic Development." World Review of Political Economy, vol. 7, no. 4, 2016, pp. 424–50. JSTOR, https://doi.org/10.13169/worlrevipoliecon.7.4.0424.

Willett, Thomas D. "The United States and World Stagflation: The Export and Import of Inflationary Pressures." The Annals of the American Academy of Political and Social Science, vol. 460, 1982, pp. 38–44. JSTOR, https://www.jstor.org/stable/1044595.

Comparing AI and Human Decision-Making in Home Loan Interest Rate Predictions
By Charishma Konda

Abstract

This study examines the role of AI in financial decision-making, with the research regarding loan applications. This research is meant to compare AI and human decision making, especially focusing on which factors contribute to the accuracy of the model. In this context, accuracy is defined as the proximity in decision from the AI model to the human decision; the closer the AI prediction is, the more accurate it will be. Using a data set of nearly 10,000 home loan records from 2017, the AI model was trained using 2,500 of the applicants, and told to predict the interest rates for the remaining loans. The results show that the model was more accurate for applications with higher annual income and larger loan amounts. These findings suggest that while AI can be used to replicate some aspects of decision making, it may also weigh certain aspects more heavily and reinforce quantitative bias present in historical lending data.

Introduction

This study examines how a trained Artificial Intelligence bot will judge loan decisions, in comparison to human employees at lending banks. The research question we will be answering is: Do AI-driven loan decisions improve accuracy or simply reinforce bias through emphasis on certain factors? Such bias would entail personal decisions made by human lenders based on their own beliefs and past experiences. Our results show us that for the most part, there are certain factors provided to lenders that also weigh more heavily into the AI model's decisions than alternative variables.

This study investigates whether AI-driven loan decisions differ from human lenders' decisions in how they weigh financial variables when determining home-loan interest rates. It is hypothesized that the AI model will place greater emphasis on quantitative variables such as annual income and debt-related measures than human lenders. Though qualitative variables were initially provided, our study is focused on quantitative measures only, as they are more easily measured and less open to interpretation.

Literature Review

The point of this study is not to observe the accuracy of the AI, but to instead observe for any bias that may be present, and see how heavily it weighs on certain aspects in comparison to humans. Researchers have increasingly examined whether AI-based decision systems can provide fairer outcomes than human decision-makers in financial contexts. Studies have been conducted with the intention to "uncover underlying sources, evolution dynamics, and associated impacts of bias" (Hu et al.). Although AI systems are often implemented with the purpose of reducing human bias, researchers have found that models may still reproduce this same bias, depending on how they were trained. Addressing challenges based on bias

necessitates certain strategies, such as "algorithmic fairness techniques, rigorous model auditing, and diversified training data" (Dubey and Vachher). Research has shown that AI systems often learn from given datasets containing human decisions, which can allow existing social biases to be replicated within algorithmic systems (Barocas and Selbst).

Machine learning systems are often trained on datasets that reflect past human decisions and social patterns. "Research in machine learning (ML) has argued that models trained on incomplete or biased datasets can lead to discriminatory outputs" (Miceli et al.). This study is meant to demonstrate how certain demographic groups may still be at a disadvantage, even without human bias in the way, with a focus on decision-making in banking and lending contexts. Traditional lending decisions often combine quantitative financial data with subjective human bias, leading to inconsistency and discriminatory behavior during loan approval. Minority applicants "are less likely than white applicants to receive algorithmic approval from race-blind government automated underwriting systems (AUS)" (Bhutta et al.). While at times, this may allow for flexibility and empathy in decision making, it can also lead to inconsistency and subjective bias.

Researchers have also compared the transparency of AI systems versus humans. Studies have been conducted, as "automated decision-making is being held to an unrealistically high standard, possibly owing to an unrealistically high estimate of the degree of transparency attainable from human decision-makers" (Zerilli et al.). While human decisions can be explained, AI models, especially ones with high levels of complexity, can make it difficult to understand how a final decision was reached. With the rise of AI, "majority of these models are inherently complex and lack explanations of the decision making process" (Hassija et al.). However, by including graphs concentrated on certain factors, we could get an idea of which parts more heavily impact the outcome. Although previous studies have examined algorithmic bias in financial settings, fewer studies directly compare how AI models weigh demographic and financial variables relative to human decision-makers in loan approval context. This study aims to address that gap by comparing patterns of decision-making between humans and trained AI models to determine whether AI systems reduce or encourage bias.

Data and Methodology

The current research utilizes data from a 2017 home loan database with nearly 10,000 entries. Relevant data include home loan applicants job titles, annual incomes, loan request amounts, total credit limits, and loan interest rates. Approximately 200 rows of data were removed because they lack job titles (a data category necessary for the current research).

We developed and trained an AI model using data from the first 2,500 entries. The AI model developed was titled 'Simple ML for Sheets', in which we used the learning algorithm called 'Gradient Boosted Trees'. The purpose of this was to help the AI understand patterns in the data, and use the skills it learned to predict new ones as well. After training our model, we inputted the remainder of the data and let the model determine interest rates. Once interest rates were predicted, we used Google Sheets features to calculate and list the differences between the

true values and the predicted values. Finally, we created graphs comparing the accuracy of the predictions between the difference in interest rates with other decision making factors inputted during the training of the AI. The purpose of this was to see which variables most heavily impacted the AI's decision making.

Independent variables of the data set include annual income, debt to income rate, credit limit, and loan amount requested. The primary dependent variable of the study was the accuracy of the interest rate prediction. Accuracy was measured as the difference between the AI-predicted interest rate and lender-assigned interest rate.

Results

The AI model predicted interest rates more accurately for high-income loan applicants. All applicants were split at the average income level, with higher than average being classified at high-income and lower than average being low-income. More accurate predictions were defined as smaller absolute differences between AI-predicted and lender-assigned interest rates.

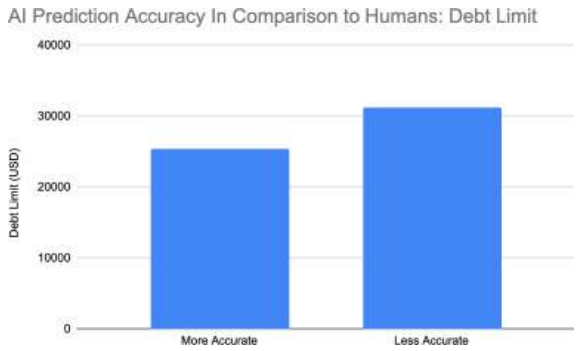


Figure 1. Prediction Accuracy by Debt Limit

There is a substantial difference between the accuracy of the debt limit chart, with the majority of the AI predictions being less accurate to the human made. Debt limit is defined as the maximum amount a lender will allow you to borrow against your property.

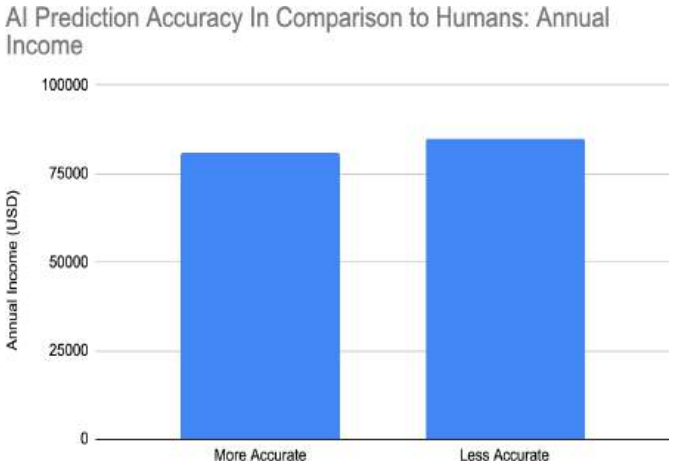


Figure 2. Prediction Accuracy by Annual Income and depicts a generally even distribution between the more and less accurate groups. Annual Income is how much one earns in a year, after taxes are cut.

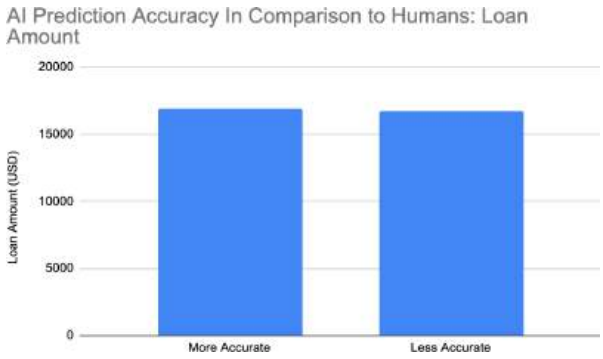


Figure 3. Prediction Accuracy by Loan Amount and contains an almost exactly even distribution between more and less accurate predictions. Loan amount did not appear to substantially influence prediction accuracy. Loan amount is the requested home equity loan amount in USD.

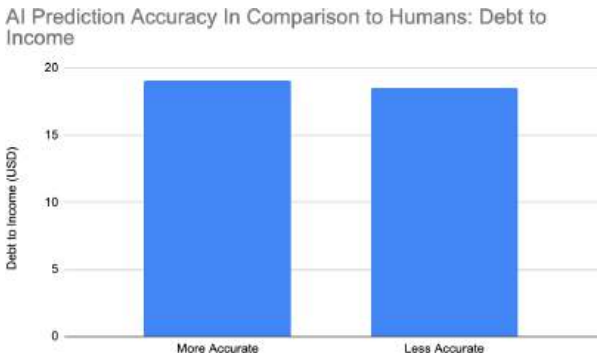


Figure 4. Prediction Accuracy by Debt to Limit Income and contains a slightly higher amount of more accurate predictions for debt to income ratio. Debt to income ratio measures the percentage of your gross monthly income spent on recurring debt payments (McMillin).

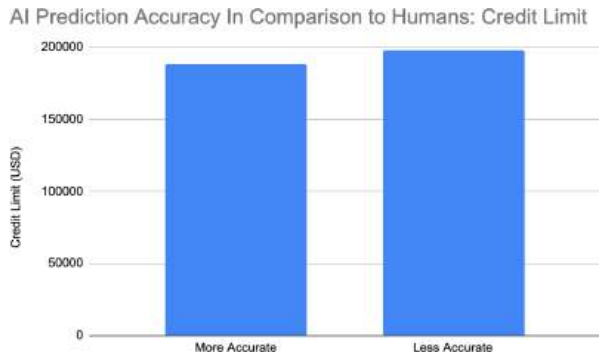


Figure 5. Prediction Accuracy by Credit Limit

As observed by the majority of the previous graphs, Figure 5 does not present a significant difference in the distribution of more versus less accuracy of the AI predictions. However, unlike other graphs, this figure shows us that there is a slightly higher amount of less accurate predictions made by models. Credit limit refers to the maximum amount of money a lender lets you borrow.

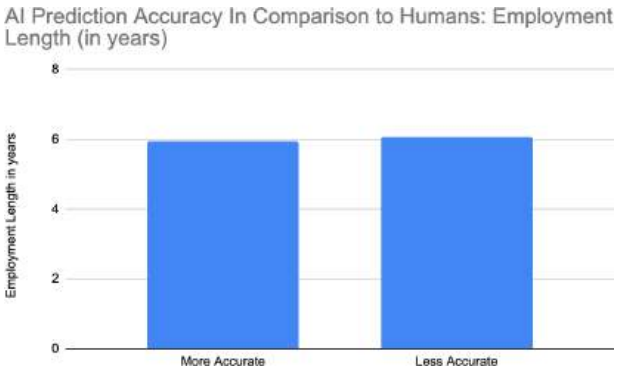


Figure 6. Prediction Accuracy by Employment Length and is again almost exactly identical in terms of accuracy distribution. Employment length signals the amount of time in years an individual has worked for a specific organization, complete with an official start and end date.

Analysis

The current research compares an AI model's estimation of interest rates for home loan applicants in comparison with human interest rate calculation from the original lender data. It then employs correlation analysis to determine which variables influenced the model's outcomes. Results show that certain variables influenced the model's output more than others. The most notable example of this was the annual income of applicants. Figure 2 shows the AI model's lesser accuracy in predicting the lender's interest rate for applicants with higher debt-to-income ratios. Employment length is not correlated with the model's prediction

accuracy for interest rate.

AI model predictions of human-calculated interest rates showed some similarities with the human calculations, but the model's outputs appear to have been influenced more by annual income than were the human calculations. A limitation of this experiment was the reasoning behind why the AI model was able to yield more human-like results in the face of higher y-value input. In addition, we also are unable to account for the thought process of the humans behind this decision making, and any bias they may have had when making their decisions.

Overall, our experiment is able to help the future by providing companies with a real experiment on the accuracy of AI in comparison to humans and allowing them to decide for themselves what they may wish to do with this information. More specifically, this research is helpful to companies who may use AI-made data years into loans to determine whether human employees have effectively served the lending company as well as an AI could have.

Conclusion

Based on the outcomes of this data, results suggest that the AI model does weigh more on certain factors than others. The AI model appears to rely more heavily on quantitative financial indicators than human lenders. These findings suggest that AI systems may reinforce existing financial bias present in human decisions if trained primarily on historical lending data. Such bias would include preference by human lenders towards applicants based on certain factors, replicated by AI to produce results more accurate to the decisions of the human lenders. While we do not know what would have been going through a person's mind as they made their loan decisions, we do know the outcome. By using the AI outcome, we can see that it is more accurate to the human decisions when considering each of the factors tested. We observe that the higher the value of the variable, the more accurate the data appears to be. Future studies should examine additional loan datasets, demographic variables, and alternative AI models. These findings are significant because institutions are increasingly incorporating AI into lending decisions, raising concerns regarding fairness and transparency. As AI systems continue to expand, understanding how these models evaluate applicants will become increasingly important for ensuring fairness, accountability, and transparency in lending decisions.

Works Cited

Barocas, Solon, and Andrew D. Selbst. “Big Data’s Disparate Impact.” SSRN Scholarly Paper no. 2477899, Social Science Research Network, 2016. https://doi.org/10.2139/ssrn.2477899.

Bhutta, Neil, et al. “How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions.” The Journal of Finance, vol. 80, no. 3, 2025, pp. 1463–96. https://doi.org/10.1111/jofi.13444.

Dubey, Shyamali, and Laxmi Vachher. “Stride Towards Fair AI: Addressing Algorithmic Prejudices in Human Resource Analytics.” 2025, p. 09702385251389336. SAGE Journals, https://doi.org/10.1177/09702385251389336.

Hassija, Vikas, et al. “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence.” Cognitive Computation, vol. 16, no. 1, Jan. 2024, pp. 45–74. https://doi.org/10.1007/s12559-023-10179-8.

Hu, Xiyang, et al. “Human–Algorithmic Bias: Source, Evolution, and Impact.” Management Science, vol. 72, no. 1, Jan. 2026, pp. 495–514. https://doi.org/10.1287/mnsc.2022.03862.

McMillin, David. “What Is A Debt-To-Income Ratio For A Mortgage?” Bankrate, 5 Sept. 2025, Accessed June 30, 2026. https://www.bankrate.com/mortgages/why-debt-to-income-matters-in-mortgages/.

Miceli, Milagros, et al. “Studying Up Machine Learning Data.” Proceedings of the ACM on Human-Computer Interaction, vol. 6, no. GROUP, Jan. 2022, pp. 1–14. https://doi.org/10.1145/3492853.

Zerilli, John, et al. “Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?” Philosophy & Technology, vol. 32, no. 4, Dec. 2019, pp. 661–83. https://doi.org/10.1007/s13347-018-0330-6.

Evaluating the Generalizability of an AI-Based Patient-Trial Matching Model in Glioblastoma By Aaron Vinoth

Abstract

Clinical trial enrollment remains one of the greatest barriers in cancer therapy. The purpose of the study is to determine the viability of an SVM model, also known as an AI model, to pair glioblastoma patients with suitable trials. This leukemia-focused model will be tested for its generalizability in glioblastoma. The model took 0.83 seconds to make 750 pairs using 15 patients and 50 trials. The cross-disease transfer of the system decreased the accuracy from 75% to 35.6%, with 0% of the trials being ranked in the Top-1. The model failed to provide a personalized list of trials for each patient. From the results, it can be concluded that an AI model cannot be implemented for diseases apart from the one for which it was designed. Understanding the findings of this study proves that AI models trained on a single disease cannot be assumed to generalize to other cancers.

Introduction

Current drug discovery in cancer is extremely slow and expensive because clinical trials have limited enrollment, long timelines, and stringent eligibility criteria (Gevorgyan et al., 2025; Kaskovich et al., 2023). Clinical trials are the gateway to new treatments, which enable more patients to obtain a cure for their disease. Glioblastoma (GBM)—a cancerous grade 4 tumor that diffusively invades regions of the brain—is a rare diagnosis and is aggressive in its progression (Fritz et al., 2016). It is extremely fatal because of its ability to embed itself into the spinal cord and brain, decreasing the ability for it to be properly removed. It has a median survival rate of around 15 months, and many therapies are ineffective in treating it because of its biological complexity as well as the limited number of patients who enroll in trials due to the rarity of the disease (Ma et al., 2021). This short timeframe for survival highlights the importance of accelerating the speed of clinical trials to increase the number of possible effective therapies.

Overall, glioblastoma trials enroll only 8-11% of newly diagnosed patients, resulting in a lack of trials advancing from phase 2 to phase 3 (Vanderbeek et al., 2018; Bagley et al., 2022). For a drug to be widely administered in the market, trials must be able to proceed to phase 3, and without this advancement, glioblastoma treatment remains limited.

A contributing factor to the limited number of therapies extending beyond the difficulty of treating the cancer is the ability to enroll patients in the trials. The eligibility criteria are non-standardized and extremely complex, which makes matching patients manually extremely time-intensive (Kaskovich et al., 2023; Mäkitie et al., 2023). The specific eligibility criteria decrease the ability for a patient to be enrolled in the “right” trial. This prevents treatment for the patient and also doesn’t afford each trial the proper number of patients required for it to pursue its project.

Currently, patient-trial matching is done manually, contributing to 86% of trials missing timelines for enrollment (Kaskovich et al., 2023; Mäkitie et al., 2023). There are vast amounts of

trials across the globe, which means a team of researchers lacks the capacity to sort through the options efficiently to match patients. However, artificial intelligence has shown that it can analyze clinical datasets, surpassing the skills of clinicians. AI has been proven to increase the effectiveness of patient matching for other cancers (Tiwari et al., 2025; Nagendran et al., 2020). This is because AI models' access to a vast range of data results in better synthesis of trial requirements, leading to better patient matching.

The research this paper focuses on revolves around testing a patient-trial matching model for glioblastoma. Analyzing this model allows for an understanding of its implementation as a pre-intervention phase for digital twins, which are virtual representations of natural, engineered, or social systems. The pre-intervention phase serves as a part of the digital twin model that can analyze patient diagnosis and trial criteria to provide treatments that are most suitable for each patient; implementing this may provide a solution to the limited glioblastoma trial enrollment. Although not currently implemented in healthcare, digital twins will allow for virtual representations of patient biology to simulate treatment outcomes (Kaskovich et al., 2023; National Academies of Sciences, Engineering, and Medicine, 2024). Implementing this model will provide substantial information for clinicians to administer the most beneficial treatment for patients, further ensuring treatment viability. In order for advancement to simulate outcomes with digital twins, patient-matching models need to be implemented first.

Artificial intelligence and patient matching systems can be the stepping stone towards implementing digital twins in the healthcare field. The digital twins framework requires that patient-specific models be used to test therapies through a simulation before actually treating the patient (National Academies of Sciences, Engineering, and Medicine, 2024). However, for that to take place, there still needs to be a patient-trial matching system to facilitate the process. Not only will the implementation of AI patient-matching systems allow for a decrease in the time required to enroll patients into trials, but it will also increase the possibility of effective treatments. Overall, an artificial intelligence model should decrease the time required for enrollment, which will not only improve treatment options but also contribute to higher survival rates for individuals. Without testing generalizability, there is an increased risk of AI models providing misleading recommendations in the clinical environment.

Literature Review

Current Climate

Research regarding AI implementation in clinical settings shows how it has strong diagnostic, prognostic, and treatment planning potential in cancer therapy (Tiwari et al., 2025). This shows how AI has the potential to surpass physicians in classifying patients and pairing them with a trial suited for them. However, the relatively new field of AI in healthcare requires stronger evidence to support its potential implementation. AI can analyze tumors, classify molecular subtypes, and predict survival, eventually leading to a method to simulate therapy responses. To make digital twin models effective for predicting AI therapy outcomes, accurate

trial matching should be ensured first as a foundational step in the research process.

There is a shift in oncology towards targeted therapies and immunotherapies guided by immune biomarkers, which allows for AI and digital twin implementation to facilitate the shift toward more personalized therapy (Sonkin et al., 2024). AI can serve as a support system but not a replacement for clinicians, proving how it can be a pre-intervention phase rather than an autonomous decision-maker (Tiwari et al., 2025; Mäkitie et al., 2023). This allows for a more patient-centric approach to the process that also decreases the resources needed in clinical trials, in turn allowing for more effective therapies. Adding on to that, it will then provide a way to transition into further research regarding digital twins.

A 2023 trial was able to create a patient-centric matching model for leukemia. The Support Vector Machine Model (SVM), an AI machine learning model used for classification, used by Kaskovich for leukemia, can generate matches in seconds. It classifies eligibility criteria into clinical categories and generates a ranked trial list. The system can then produce Top-1 and Top-3 matches for synthetic patients. This leukemia-focused model will be tested for its generalizability in glioblastoma. Furthermore, for digital twins to be used in the future, there needs to be a pre-intervention phase before the simulation of therapies (Shen et al., 2025). Should the pre-interphase fail to make clinically viable decisions, future paths in the digital twins framework may suffer in their performance. As a result, patient and trial outcomes would greatly suffer, further emphasizing the need for a reliable AI model for the pre-intervention phase.

Gap

Digital twins could reduce costs and ethical burdens, but there are gaps in clinical integration (National Academies of Sciences, Engineering, and Medicine, 2024). The SVM trial-matching model has only been used for pediatric leukemia. It has not been used on glioblastoma patients with GBM eligibility criteria (Kaskovich et al., 2023). Digital twins have also had early successes in breast and lung cancer, but have thus far not provided evidence of any success or trials for a patient-matching model for GBM (National Academies of Sciences, Engineering, and Medicine, 2024). AI implementation cannot rely on generalizability, as it may underperform or not perform to the same extent on a different specialized disease like leukemia vs. GBM (Reddy et al., 2021). Because of this, it provides a gap for research. The researcher could test the same SVM model created by Kaskovich et al. to see if it is effective on glioblastoma trials, a cancer that has completely different eligibility criteria. This opens up a research question to be studied: to what extent can an AI patient-matching model serve as a pre-intervention phase for digital twins in glioblastoma patients when diagnosing cancer therapy? It is important to test whether the model used for leukemia can be reused for glioblastoma without any additional adjustments. Glioblastoma has different eligibility criteria, chronic failures, and symptoms from leukemia. This is vital to test since digital twins are advertised as the future of oncology. Because of this, it should be studied whether the pre-intervention phase can work with GBM while remaining feasible.

Intro to Methodology

In order to properly test generalizability, this study follows the same methodology as Kaskovich et al., as it is also used in their model, with their permission. The experimental methodology that was used utilized public datasets, like ClinicalTrials.gov, to get the trial data and eligibility criteria. The "patients" used are a synthetic patient cohort generated from published eligibility distributions. This study will run the SVM model by feeding the eligibility sentence and the synthetic patient cohort to create a ranked trial list with composite scores for each patient. Timing will also be measured. Overall, I hypothesize that an AI model will effectively be able to pair patients with trials. However, due to the implication that this study will be using a model originally tested on leukemia, glioblastoma outcomes may be less effective than outcomes detailed by Kaskovich et al.

Methods

Design of the Study

The purpose of the study is to determine the viability of an SVM model, also known as an AI model, to pair glioblastoma patients with suitable trials. To that end, this paper will model the procedure after Kaskovich et al., a 2023 study done to determine whether an AI model can serve as a patient-centric matching model for pediatric leukemia.

To effectively replicate the methodology done by the 2023 trial, an email was sent out to the researchers of the Kaskovich et al. paper to gain permission to use their pre-trained model. Once permission was obtained, the SVM model linked to their research paper was accessed. The researchers further granted permission to use their resources and tools that were also publicly available, ensuring this study could be closely replicated to effectively compare the outcomes of glioblastoma patients with pediatric leukemia patients. This access helped determine the viability of an AI model to be generalized for the use of glioblastoma.

Design Viability

This study uses an experimental design that outputs a quantitative analysis of patient matching for glioblastoma. This is the only methodology that effectively allows for an evaluation of the performance outcomes. This paper must follow the same procedure as Kaskovich et al. because it uses the same model. Doing so eliminates the variability in the outcomes due to discrepancies in the method. A qualitative analysis would be inappropriate as there is no previous research on digital twins for glioblastoma, eliminating any methodologies related to a content analysis. Qualitative methods also prove unviable because they lack the ability to provide a percentage accuracy to determine whether an AI model can work in relation to a manual-rules baseline, which a quantitative method allows for. Ensuring an experimental design allows me to obtain the trials that each patient can be matched to once the SVM model is given the eligibility criteria, along with the patients' diagnoses. This then allows for a comparative analysis of the manual-rules baseline, the human-matched output, and the SVM model-matched data.

Ethicality

The study used a synthetic patient cohort consisting of 15 patients with varying levels of diagnosis and demographics. All data related to patients was synthetically created. The patient cohort utilized was provided by the researchers of "Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia." This keeps the procedure as similar to the model paper as possible, keeping the variability of outcomes because of the methods to a minimum. Also, using the AI-Based Support Vector Machine Model means that the tool was run and then output data of which trials each synthetic patient got matched to. An analysis was then run to determine whether the tool was effective in its operation. The use of the AI tool doesn't consist of artificial intelligence helping me to carry out the research. Instead, it is the item being inherently studied. All aforementioned parts regarding the synthetic patient cohort and the AI tool—the SVM Model—were approved by the Institutional Review Board.

Trial Selection

The ClinicalTrials.gov database was utilized to obtain trials with eligibility criteria for this study. The word "glioblastoma" was used to narrow the search. Afterwards, the recruiting status was narrowed to "recruiting OR active, not recruiting." After this, 50 trials found in the past year—December 2024 to January 2026—were downloaded. All narrowing criteria were implemented to ensure that trials used are current or ongoing. Due to time constraints, only 50 trials were extracted in comparison to the Kaskovich study, which had 256 trials. Though the number of trials used was less than the model study, 50 trials with 15 patients allow for 750 possible matches. The 750 comparisons provide for an effective analysis, even though it is a significant decrease from Kaskovich et al.'s study.

Eligibility Criteria Extraction & Discretization

After the trials were downloaded, the eligibility criteria were extracted from each trial. The individual characters of each trial's criteria section ranged from 500 to 5,000 characters. The eligibility text was then broken down into individual sentences to separate the number of different criteria each consisted of. Any sentence less than 10 characters was removed as per Kaskovich et al., being considered an "incomplete sentence." Following this, 708 individual sentences (criteria) were found across the 50 trials, giving 14.2 criteria per trial.

SVM Classifier Loading

Kaskovich linked their pre-trained models on their GitHub storage, a cloud server that allows for the storage and sharing of coding projects, which housed 17 category-specific SVM classifiers. The classifiers work by taking a criterion (eligibility sentence) and scoring whether it can be matched with its category by giving the criterion a score of one or zero. Each criteria matching is then given a confidence interval. Once each criteria has been run on all 17 SVM classifiers, the category with the highest confidence score is selected.

Word Embedding

Word embedding is the process by which each word is converted into a number: each word is a number, each sentence is a vector, and the vector length is a dimension. In the Kaskovich paper, each sentence was given 256 numbers (256-dimensional embeddings) that were generated by using FastText (an open-source library used for efficient text classification). The 256-dimensional sentence embeddings were trained on medical text.

The FastText model that Kaskovich used lacked certain components when it was used for this study, rendering the model incomplete and unusable. To overcome this, two options were considered: download a pre-trained biomedical embedding model or use a general-domain embedding. The general-domain embedding was ultimately the most viable option. Because the biomedical embeddings are trained on PubMed abstracts, full-text journal articles, or clinical notes, they are often behind paywalls restricted by licenses and governed by institutional agreements, making them an unfeasible resource for this study. However, general-domain embeddings are publicly available and require no credentials, allowing for a feasible option to embed words into vectors. Due to these factors, I implemented GloVe (Global Vectors for Word Representation) embeddings trained on Wikipedia and Common Crawl, which is the most accessible option for a general-domain embedding.

Dimension Adjustment

The downloaded SVM was trained on embeddings that consisted of 256-dimensional input vectors, but the GloVe only provided 200-dimensional vectors. To compensate, 56 zeros were added to each vector, keeping the information consistent.

Synthetic Patient Cohort

Each patient was assigned demographics such as age, gender, and performance status using the Karnofsky scale. In addition, vitals related to glioblastoma, including blood count, renal function, diagnosis type (newly diagnosed, recurrent, progressive), and treatment history (chemotherapy, radiation, and surgery) were assigned.

As a result of the synthetic cohort being created for leukemia, the diagnosis for each patient had to be modified to replace leukemia with glioblastoma under the patient description. Random ranges for GBM epidemiology were created and substituted, such as 60% having recurrent GBM.

Patient-Trial Matching

Finally, the study conducted a rule-based matching to determine whether each eligibility criteria fit the criteria labeled by the trials. For every trial a patient was matched to, a match score of the percentage of trial eligibility was calculated. The 15 patients were all matched to every trial of the 50 presented to get 750 comparisons. Then, all the trials were ranked with a match score, along with measuring the processing time for the system. The match score was created by

taking the proportion of matches to potential matches, which were then added and divided by the potential matches to get a match score.

Manual-Rules Baseline

To create a baseline that acts similarly to a real doctor to compare the patient-trial matching model, an API (Application Programming Interface) key is fed with the same data as the SVM model. The key generates an alphanumeric code that needs to be authorized to carry out the function. Once run, it gave data detailing whether the patient was eligible or not for a trial, where a patient fell short of being accepted, and how confident it is in the decision.

Limitations

The general domain used for the word embedding lacked the specialized knowledge that is present in a biomedical domain. This might have resulted in some of the eligibility criteria not being accurately transformed into numbers that the SVM model must read. Furthermore, the API key that was meant to create a substitute for a doctor matching the patients was only able to match 15 of the 20 original patients used by Kaskovich et al. As a result, instead of having 1,000 comparisons, this study was only able to evaluate 750. This was because the key "went public," keeping it from continuing its function. To rerun it would have been too expensive. As a result, the comparison between the SVM model and the manual-rules baseline lacked the ability to give an accurate comparison for 5 patients; it also deprived the study of 250 possible matches to compare.

Data Results

Classification Results

Category	**Count**	**Percentage**
Concurrent Medications	361	51%
Blood Counts	301	42.5%
Active Infection	3	0.4%
Marrow Blast Percentage	1	0.1%
Unclassified	42	5.9%
All Other Categories	0	0%
Total	708	100%

Figure 1

In Figure 1, the SVM Model matched all patients and trials after identifying 5 categories. Of the 708 criteria, Concurrent Medications were 361 (51%), Blood Count was 301 (42.5%), Active Infection was 3 (0.4%), and Marrow Blast Percentage was 1 (0.1%). The other 12 categories were unable to be detected by the model.

SVM Matching

Metric	Value
Total Comparisons	750
Processing Time	0.83 seconds
Average Match Score	93.6%
Median Match Score	100%
Pairs Scoring 90-100%	476 (63.5%)
Pairs Scoring 70-89%	266 (35.5%)
Pairs Scoring Below 70%	8 (1.1%)
Unique Top-5 Sets	1 (all patients had identical matches)

Figure 2

The model took 0.83 seconds to make 750 pairs using 15 patients and 50 trials. The time was measured and provided by the SVM model after it had completed its analysis. It generated an average match score of 93.6% with a median of 100%. The distribution shows that 476 pairs (63.5%) scored between 90% and 100%, 266 pairs (35.5%) scored between 70% and 89%, and 8 pairs (1.1%) scored below 70%. Each patient received the same top-5 recommendation, meaning that the same trials were ranked in the same order.

Match Scores by Eligibility Status

Manual Decision	SVM 90-100%	SVM 70-89%	SVM Below 70%	Total
Eligible	5 (83.3%)	1 (16.7%)	0 (0%)	6
Ineligible	395 (61.6%)	238 (37.1%)	8 (1.2%)	641
Uncertain	76 (73.8%)	27 (26.2%)	0 (0%)	103
Total	476	266	8	750

Figure 3

Among the eligible pairs, 5 were given a score of 90-100% by the SVM, and one was

given a score of 70-89%. For trials classified as ineligible, the SVM provided a high match score of 90-100% for 395 pairs, 70-89% for 238 pairs, and only 8 received a score below 70%. Although the SVM classified 103 pairs as uncertain, it assigned scores of 90-100% to 76 pairs, with only 27 receiving scores of 70-89%.

Top-K Accuracy

Metric	Accuracy	Successes
Top-1	0%	0/6
Top-3	16.7%	1/6
Top-5	16.7%	1/6
Top-10	33.3%	2/6

Figure 4

The Top-1 accuracy resulted in 0% after the SVM model produced 0 trials as the most highly recommended for a patient classified as eligible based on the manual baseline. It had a Top-3 and Top-5 accuracy of 16.7% after ordering 1 of the 6 trials classified as a suitable fit by the manual baseline. Looking at the Top-10 accuracy, the model had an accuracy of 33.3% after putting 2 of the 6 eligible in the top 10 rank. Four of the eligible trials classified by the baseline failed to be highly ranked by the SVM model, all lying outside the top-10 list.

Eligibility Trial Ranking Detail

Patient	Eligible Trial	SVM Rank	SVM Score
GBM Patient 3	NCT00068003	34	75%
GBM Patient 5	NCT03970447	13	100%
GBM Patient 7	NCT01044966	2	100%
GBM Patient 11	NCT00879427	18	100%
GBM Patient 13	NCT03970447	13	100%
GBM Patient 14	NCT00052013	7	100%

Figure 5

The manual-rules baseline provided an output of 6 eligible trials for 6 of the patients, compared to the SVM model, which only ranked two of the trials in the top-10. The eligible

trials were ranked as follows for each patient: patient 3, ranked 34; patient 5, ranked 13; patient 7, ranked 2; patient 11, ranked 18; patient 13, ranked 13; patient 14, ranked 7.

Analysis and Discussion

Analysis Choice

A descriptive analysis and a comparative analysis are used to understand the validity of the SVM model for patient-trial matching. Match scores, trial ranking, and accuracy of the model to the manual-rules baseline were used to compare the results with Kaskovich et al.'s performance.

Performance on GBM

The SVM model had an overall confidence score of 35.6%, a significant decrease in comparison to Kaskovich's 75% accuracy. Although the SVM model used was the same, the performance of the model was not transferable from leukemia to glioblastoma. This shows how the model cannot be universally implemented across cancers.

The SVM model trained on pediatric leukemia consisted of 17 categories that needed to be analyzed for each patient. However, there is a vast difference in the language used for eligibility criteria between glioblastoma and leukemia. Because the model was previously used for leukemia, it lacked the ability to understand the different eligibility criteria and diagnosis used for glioblastoma, which led to misclassifications and inflated match scores. As a result, the model could only recognize five of the seventeen categories it needed to classify to give accurate responses for each patient. This is only a 29% coverage, unlike Kaskovich's 100% coverage. The 12 undetected categories often used in GBM trials were unable to be classified.

Kaskovich et al.'s pairing used all the categories, and the categories were equally used to match trials to patients. However, this study found that 2 categories, Concurrent Medications and Blood Counts, were responsible for 93.5% of all the matches. This data is extremely skewed, resulting in an unequal representation of the trials. This led to misclassifications across all comparisons.

Matching Performance

Kaskovich et al. did not specifically state their Top-K analysis. However, they stated their experiment resulted in a 75% accuracy, implying the superior matching capability of the model on leukemia in comparison to GBM. In this study, the model had an output of 0% accuracy in the Top-1 Accuracy. This means it failed to rank the eligible trials as the top preferences, which means it failed to accurately carry out its function.

Although the accuracy of the results is limited, the system took 0.83 seconds to complete 750 comparisons. This is a significant advantage in comparison to manual matching, which, on average, can take 400 hours. Kaskovich et al. stated that their experiment took 3-4 seconds to carry out its function. Both of these tests exemplified the speed of the model. However, this

study had no accuracy in association with its speed, as seen with its 0% top-1 accuracy.

Another important aspect that was greatly emphasized in Kaskovich et al.'s paper was that the SVM model provides a personalized trial recommendation. By comparison, this study had identical patient-trial matches for all 15 patients regardless of the clinical variables, indicating the model lacked the ability to differentiate between patients and trials, failing in a key aspect that is expected.

Score Distribution

Out of the 750 comparisons, 63.5% had a match score of 90-100%, 35.5% had a match score of 70-89%, and 1.1% scored below 70%. The SVM model had an average score of 93.6%, indicating that most of the patient matches were nearly perfect in their accuracy. To compare the accuracy given for eligible vs ineligible trials was 83.3% to 61.6%. This score inflation was caused by the failures of the model to accurately classify each patient and trial, and resulted in eligible trials with a 100% match score ranked 7th, 13th, and 18th. Many of the ineligible trials also scored 100% while being ranked higher, detailing the difficulty of the model to differentiate between eligible and ineligible matches.

Validation Significance

The manual-rules baseline allowed for the SVM model's 93.6% match score to be disproven. Comparing the results of the SVM model to the baseline proves that when testing new systems, it is vital that results are compared to expert judgment. This is because match scores and confidence intervals can be severely misleading. The high match score is a result of the inflation caused by the inaccuracy of the model. Of the 750 pairs, the baseline identified 6 eligible pairs (1% eligibility rate), which is consistent with glioblastoma trials' extreme requirements. The low rate showcases how eligible trials are unlikely and that the system was not able to identify them. This proves that the SVM model cannot be used for GBM trials because it is not clinically valuable.

Counterargument

A possible solution to address the mentioned problems could be to retrain the model for GBM data. However, this retraining would require significant monetary investment because it would involve manual labeling of thousands of criteria, computational costs, and domain expertise. The clinical community's objective should be to determine methods that allow for efficiency while eliminating unnecessary costs. This is why, rather than retraining a model for multiple cancers, the universality of models should be tested to understand factors that limit it.

Limitations of the Study

There were several limitations that may have contributed to the inability to apply the SVM model to glioblastoma. This study analyzed 50 GBM trials, which is 166 trials fewer than Kaskovich's 216 leukemia trials. Even though this was statistically lower, the sample size was

proficient at portraying the barriers that don't allow for universal use of the model. Furthermore, the manual-rules baseline was run using a Large Language Model-based Application Programming Interface to simulate a physician's judgment rather than a doctor carrying out this action. This is a common tool that is used in similar studies when experts are unavailable, but it still lacks the clinical judgment of a licensed physician. The API key, although a viable approach, was limited to only evaluating 15 out of the 20 patients that Kaskovich et al. used. This decreased the trial matches from 1000 to 750, eliminating 250 matches that could have provided more data regarding top-1 accuracy or inflated scores.

Finally, to translate the words into numbers for the SVM model to read the eligibility criteria, a general domain embedding trained on Wikipedia and web text rather than medical literature was used. Unfortunately, this lacked the accuracy of the medicine-specific embedding used by Kaskovich. This means that certain medical terms didn't have the statistically significant translation they should have had. Secondly, the embedding model was required to transform each sentence into 256 numbers regardless of the length of the sentence, according to Kaskovich. However, this model transformed all sentences into 200 numbers. This resulted in the model having to add 56 unimportant zeros that the SVM model wasn't trained to take into account.

This study portrays the lack of efficiency in implementing the same clinical trial matching system for different diseases. It implemented the Kaskovich methodology and noted that the trial matching process was being decreased to less than a second, a significant increase in efficiency. However, the cross-disease transfer of the system decreased the accuracy from 75% to 35.6%, with 0% of the trials being ranked in the Top-1. The model failed to provide a personalized list of trials for each patient, as it gave identical trial rankings even though each trial and patient had different characteristics. From the results, it can be concluded that an AI model cannot be implemented for diseases apart from the one for which it was designed.

Hypothesis Rejection and Implications

Hypothesis Rejection

Overall, the hypothesis that the AI model will effectively be able to pair patients with trials was rejected, as the results have proven that the SVM failed to provide an accurate classification of eligible results.

Implications

Understanding the findings of this study proves that AI models trained on a single disease cannot be assumed to generalize to other cancers, further expanding on the conclusion of Reddy et al. The results of this study reiterate how the implementation of an AI model in clinical settings can lead to risks if not properly validated. The SVM model's decline in performance from leukemia to GBM indicates the need for retraining, should cross-application be necessary. Furthermore, the paradox of the model's efficiency in processing speed to its inefficiency in accuracy illustrates the complexity of factors that comprise a model's possible application in a

real-world context. Regardless of its catalysis of the process, the SVM model's accuracy proves a significant area that requires improvement for clinicians to trust its eventual implementation.

In the context of the larger digital twins framework, an inaccurate pre-intervention phase would lead to consequent inaccuracies in treatment outcome predictions, reiterating the concept described by Shen et al. As a result, patients would suffer the burden of ineffective and possibly harmful treatments. Also, this would lead to a continuous cycle of GBM trials unable to advance to phase 2 and phase 3, slowing the innovation of new therapies for this highly lethal cancer. The consequent negative outcomes that can occur as a result of an inaccurate pre-intervention phase further demonstrate the need for an optimal AI model.

Future Direction and Conclusion

Future Direction

Should this study be conducted again, there are certain changes that could be implemented to obtain a more accurate comparison. For instance, the FastText model should be obtained from the Kaskovich team to have the proper embeddings, which can thus improve classification performance and decrease issues that may have resulted from embedding issues rather than fundamental issues.

For future AI models, developing a multi-disease training dataset that could span various cancers such as leukemia, glioblastoma, lung cancer, breast cancer, and other common cancers would allow the model to be used as a universal tool by learning multiple disease-specific criteria. It would also allow the model to be trained on multiple different vocabularies specific to each disease. Removing disease-specific models and creating a universal model would allow for a decrease in barriers for it to be implemented in the clinical setting as well.

Should such a model be created and tested on synthetic patients, it would have to be tested in a real hospital setting to measure actual clinical reliability. It would also be beneficial for the model to describe what criteria each patient meets and fails to meet, as it would allow clinicians to further trust the model and allow them to override decisions should they see fit.

Conclusion

Clinical trial enrollment remains one of the greatest barriers in cancer therapy. Artificial intelligence models have the potential to eliminate this barrier to allow for more personalized trials for each patient. Even though a universal model has not been created, this study provides problems that, if addressed, can pave a path towards it. To increase the probability of its creation, IT specialists and physicians need to continue to work together to increase the accuracy of AI models. Through continued collaboration, a universal model has the possibility to increase the standard of care that all patients receive.

Works Cited

Bagley, S. J., Kothari, S., Rahman, R., Lee, E. Q., Dunn, G. P., Galanis, E., Chang, S. M., Nabors, L. B., Ahluwalia, M. S., Stupp, R., Mehta, M. P., Reardon, D. A., Grossman, S. A., Sulman, E. P., Sampson, J. H., Khagi, S., Weller, M., Cloughesy, T. F., Wen, P. Y., & Khasraw, M. (2022). Glioblastoma Clinical Trials: Current Landscape and Opportunities for Improvement. Clinical cancer research : an official journal of the American Association for Cancer Research, 28(4), 594–602. https://doi.org/10.1158/1078-0432.CCR-21-2750

Fritz, L., Dirven, L., Reijneveld, J. C., Koekkoek, J. A., Stiggelbout, A. M., Pasman, H. R., & Taphoorn, M. J. (2016). Advanced Care Planning in Glioblastoma Patients. Cancers, 8(11), 102. https://doi.org/10.3390/cancers8110102

Gevorgyan, A. (2025, September 26). Artificial Intelligence in Cancer Drug Discovery in 2025 - OncoDaily. Oncodaily. Retrieved December 8, 2025, from https://oncodaily.com/oncolibrary/artificial-intelligence-in-cancer-drug-discover

Kaskovich, S., Wyatt, K. D., Oliwa, T., Graglia, L., Furner, B., Lee, J., Mayampurath, A., & Volchenboum, S. L. (2023). Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia. JCO clinical cancer informatics, 7, e2300009. https://doi.org/10.1200/CCI.23.00009

Ma, W., Wang, Y., Zhang, R., Yang, F., Zhang, D., Huang, M., Zhang, L., Dorsey, J. F., Binder, Z. A., O'Rourke, D. M., Fraietta, J. A., Gong, Y., & Fan, Y. (2021). Targeting PAK4 to reprogram the vascular microenvironment and improve CAR-T immunotherapy for glioblastoma. Nature cancer, 2(1), 83–97. https://doi.org/10.1038/s43018-020-00147-8

Mäkitie, A. A., Alabi, R. O., Ng, S. P., Takes, R. P., Robbins, K. T., Ronen, O., Shaha, A. R., Bradley, P. J., Saba, N. F., Nuyts, S., Triantafyllou, A., Piazza, C., Rinaldo, A., & Ferlito, A. (2023). Artificial Intelligence in Head and Neck Cancer: A Systematic Review of Systematic Reviews. Advances in therapy, 40(8), 3360–3380. https://doi.org/10.1007/s12325-023-02527-9

Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., Topol, E. J., Ioannidis, J. P. A., Collins, G. S., & Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ (Clinical research ed.), 368, m689. https://doi.org/10.1136/bmj.m689

National Academies of Sciences, Engineering, and Medicine; National Academy of Engineering; Division on Earth and Life Studies; Division on Engineering and Physical Sciences; Board on Atmospheric Sciences and Climate; Board on Life Sciences; Computer Science and Telecommunications Board; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Committee on Foundational Research Gaps and Future Directions for Digital Twins. (2024). Foundational Research Gaps and Future Directions for Digital Twins. National Academies Press (US). https://doi.org/10.17226/26894

Reddy, S., Rogers, W., Makinen, V. P., Coiera, E., Brown, P., Wenzel, M., Weicken, E., Ansari, S., Mathur, P., Casey, A., & Kelly, B. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ health & care informatics, 28(1), e100444. https://doi.org/10.1136/bmjhci-2021-100444

Shen, S., Qi, W., Liu, X., Zeng, J., Li, S., Zhu, X., Dong, C., Wang, B., Shi, Y., Yao, J., Wang, B., Jing, L., Cao, S., & Liang, G. (2025). From virtual to reality: innovative practices of digital twins in tumor therapy. Journal of translational medicine, 23(1), 348. https://doi.org/10.1186/s12967-025-06371-z

Sonkin, D., Thomas, A., & Teicher, B. A. (2024). Cancer treatments: Past, present, and future. Cancer genetics, 286-287, 18–24. https://doi.org/10.1016/j.cancergen.2024.06.002

Tiwari, A., Mishra, S. & Kuo, TR. Current AI technologies in cancer diagnostics and treatment. Mol Cancer 24, 159 (2025). https://doi.org/10.1186/s12943-025-02369-9

Vanderbeek, A. M., Rahman, R., Fell, G., Ventz, S., Chen, T., Redd, R., Parmigiani, G., Cloughesy, T. F., Wen, P. Y., Trippa, L., & Alexander, B. M. (2018). The clinical trials landscape for glioblastoma: is it adequate to develop new treatments?. Neuro-oncology, 20(8), 1034–1043. https://doi.org/10.1093/neuonc/noy027

A Study of Distance to Criticality in Conservative Zhang Sandpile Models of Random Neural Networks By Kean Li and Afshin Montakhab

Abstract

Research on the critical brain hypothesis centers on the computational benefits of functioning at a continuous phase transition, meaning the accurate detection of criticality from measurable activity is a requirement. Previous studies have typically focused on the analysis and fitting of power laws to size and duration distributions of neural avalanches as markers of criticality. Recently, a new method of measuring the state of such systems is the DCC coefficient $\mathrm{DCC} = \gamma - \frac{\beta-1}{\alpha-1}$, or the difference between measured and expected scaling exponents. We tested this new measure on an energy-conserving discrete-time continuous sandpile with energy transmission from firings based on an Erdos-Renyi graph, using 4 different biologically-inspired triggering methods to create activity. We collected 1 million avalanches per system size, energy, and trigger method, and fit exponents using maximum-likelihood estimation for distributions and Theil-Sen regression on log-log transformed data for γ. After accounting for finite-size-scaling-effects, we found that DCC only converges to 0 for 3 of the triggers as energy approaches the measured critical energy, and even then, DCC does not converge smoothly as previously expected. Despite this, the underlying γ seems to be a useful measure of criticality, corroborating previous literature.

1. Introduction

The brain criticality hypothesis posits that biological neural networks function at a phase transition between absorbing and active states, maximizing computational ability and sensitivity to stimulus (Beggs and Timme 163). Significant effort has been dedicated to supporting this hypothesis, mainly through the analysis and fitting of various power laws on data derived from experimental measurements of neuronal avalanches that characterize such critical behavior (Beggs and Timme 163). Traditionally, the existence of power-law distributions of the sizes, $P(s) \sim s^{-\alpha}$, and durations, $P(T) \sim T^{-\beta}$, of neural avalanches was used as evidence for criticality, per Beggs (14), but more recently, a scaling relation that measures the nonlinear relation between the average size of the avalanche as a function of its duration has attracted more attention, $\gamma \approx (\beta - 1)/(\alpha - 1)$ (Beggs and Timme 163). When this relation is not exactly satisfied, it suggests that the system is not critical (Ma et al. 657), but one might wonder if there is more information to be extracted from this difference between the measured scaling exponent γ and the predicted one from the scaling relation. In this study, we will shed light on the behavior of a possible new measure of distance to criticality, the aforementioned difference $\gamma - \frac{\beta-1}{\alpha-1}$. We will call this the distance to criticality coefficient, or DCC.

We will study DCC on a modified energy-conserving sandpile model tuned to criticality. We will use the same simple toppling (avalanche) behavior that defines regular self-organizing critical continuous sandpiles, but will run it on a complex network instead of a lattice, making it

more biologically realistic. We will use four types of stimuli to trigger avalanches in our model. Lastly, we will collect and analyze the avalanche our model produces and determine the viability of DCC as an accurate measure of a tunable system's critical state.

2. Model

2.1 Overview

To investigate the accuracy of DCC as a marker of criticality, we simulate a perfect (energy conserving) model of neurons. The state of the system is determined purely by the set of energies $E = \{E_1, E_2...E_N\}$, with a single energy describing the state of a single neuron. Neurons with an energy above 1 are considered active. Energy, and by extension the system, evolves synchronously in discrete timesteps.

2.2 Toppling behavior

In a given timestep, active neurons ($E \geq 1$) are marked to "fire". In this process, the neuron evenly distributes its existing energy to its neighbors (other neurons that it is connected to). The neuron's energy is then subtracted from itself, meaning in an isolated case, a firing neuron ends with 0 energy. When multiple neurons fire in the same timestep, the energy changes are summative. This process always conserves energy.

Formally, E_i at timestep $t + 1$ is given by

$$E_i(t + 1) = E_i(t) - E_i(t)\,\Theta\big(E_i(t) - 1\big) + \sum_{j \in N_i} \frac{E_j(t)}{K_j}\,\Theta\big(E_j(t) - 1\big)$$

where N_i is the neighbor set of i. The step function $\Theta(x)$ equals 1 when $x \geq 0$ and 0 when $x < 0$. The term $\Theta(E_j(t) - 1)$ becomes 1 if neuron i is active, causing its energy to reset, while the summation term adds incoming energy only from neighbors j that are themselves active.

In our model, an active system evolves deterministically. Firing neurons may cause their neighbors to become unstable and fire in the subsequent timestep. These cascades of firings constitute "avalanches" of activity, which are defined by their size and duration. Duration is the number of timesteps that an avalanche spans, starting from the timestep of the first firing and including the last timestep with active sites. The size of an avalanche is the total number of firing events across all timesteps. Respectively, $P(s)$ and $P(T)$ denote the distribution of avalanche sizes and durations. Neurons can fire multiple times per avalanche; therefore, avalanches can exceed the system size N.

2.3 Network setup

In our model, the connections between neurons are modeled by randomly generated $G(n, p)$ undirected Erdos Renyi graphs, where nodes are neurons, and edges are connections that transfer energy in firing events. If an edge exists between two nodes, the corresponding neurons are said to be neighbors. Self-connections are disallowed, meaning firing neurons in isolated cases end up with 0 energy. Graphs are labeled with their size N and average degree

(connectivity) K, which defines the connectivity probability p. We study graphs of size $N \in \{20000, 40000, 80000\}$ and average degree $K = 500$. The degree (number of neighbors) distribution is Poisson and at the system sizes and average degree used, all networks are above the percolation threshold.

2.4 Triggering avalanches

When the network is unstable, or there are firing neurons with energy greater or equal to 1, the system proceeds per the avalanche dynamics specified in 2.1. Because we study the model from the absorbing phase up to the phase transition to the active phase, avalanches will eventually end. When this happens, the energy of the system is manipulated through stochastic energy-conserving perturbations. It is fundamental that the total energy remains constant so we can define a precise energy and characterize the system at that particular energy. Consequently, we have defined four different trigger methods which we believe have important implications for self-organized critical systems. We call these methods: Force-fire, transfer, delayed-transfer, and addition.

Force-fire involves randomly selecting a neuron in the system and forcing it to fire, regardless of whether its energy is above threshold; this method replicates the random firings known to happen in the brain (Burns and Webb 211). Force-fire is a uniform addition of energy to one neuron's neighbors, with energy conservation enforced through an equal subtraction from the original neuron. No negative energies are possible or allowed.

The addition method is very different. It involves selecting a random neuron, adding a random "bump" of energy $\delta E \sim U[0, 0.2]$, and evenly subtracting an amount $\frac{\delta E}{N-1}$ from all other neurons. This method can create negative energies. While negative energies have no direct physical interpretation in the context of energy representing membrane potential, it may serve to simulate the effects of a refractory period, inhibitory neurons, or global depressant. Overall, the addition method simulates external stimulus.

However, the addition method produced vastly different results from the force-fire method when it was first investigated. Therefore, we added the transfer method, meant to be a hybrid of the two. Transfer involves the same addition of δE to a random neuron, but instead of globally subtracting to ensure energy conservation, another random neuron is chosen to have δE subtracted from it. This also can create negative energies. This may also simulate external stimulus.

Lastly, delayed-transfer is a modified version of transfer. It involves delaying the energy subtraction. If the addition of δE causes an instability, in delayed-transfer, the avalanche is immediately allowed to happen; only at the end of the avalanche is the subtraction realized. When the addition of δE causes no instability, delayed-transfer is functionally equivalent to transfer. Comparing results from delayed-transfer and transfer may give insight to what specific microscopic changes cause the different macroscopic behavior between force-fire and addition methods.

3. Simulation methodology

3.1Data collection

We collected data for each of the trigger methods, system sizes, and average energies using a multi-threaded computer simulation. Individual threads are statistically independent, but run under the same conditions. Data is pooled across threads. For any thread, the high-level simulation process is to first initiate the energy array, then simulate avalanches, and lastly consolidate data once the target number of avalanches is reached. For each specific (network, method, energy) tuple, 1 million avalanches were collected. Before data collection starts but after energy initialization, the simulation runs for a number of "pokes", or applications of triggers to an inactive system. This ensures that our data collected is from the equilibrium state of the system.

Avalanches that occurred before the thermalization cutoff are discarded from the final analysis. Any simulation that encounters an avalanche with a size larger than $10000N$ is reset; the long-running avalanche is discarded from the data. This reset process ensures computational feasibility near the critical point, and provides an estimator to whether activity is infinite or not. While the cutoff factor is arbitrary, empirically we observed finite-size effects on avalanche size gaining significance around $s \approx N$. This leads us to conclude minimal impact from the cutoff. It is highly unlikely that avalanches larger than the cutoff terminate.

3.2 Data processing

Once data from a single (network, trigger method, energy) tuple is collected, it is processed in multiple steps to ensure reliability and obtain relevant statistics.

3.2.1 Thermalization analysis

First, data is checked for sufficient thermalization, meaning data is from a system at equilibrium. This is verified by calculating an enhanced version of the Gelman-Rubin (GR) statistic split-R, originally defined in Gelman and Rubin (457-472), on avalanche size and duration.

The GR statistic R compares within-chain variance to between-chain variance of Markov Chain Monte-Carlo algorithms, which includes our multi-threaded simulation. Variation of avalanches produced by a single thread is compared to variation of avalanches between threads. If they are sufficiently similar, the system is considered to be at equilibrium. $R \leq 1.02$ is used as our cutoff (Vats and Knudson 518-529).One of the requirements of GR is sufficient starting variation between threads. To ensure this, the energy array is initialized according to a Bernoulli distribution, with possible energies 0 and $2E_{avg}$. Half of the energies have each value, giving an average of E_{avg} and a variance of E_{avg}.

To ensure compatibility with the expected power-law distributions of avalanche size and duration, we compare our effective R to the cutoff. Effective R is the maximum of rank-normalized split-R and folded split-R, described by Vehtari et al. (667-718). We use lugsail estimators for within-chain variation, per Vats and Knudson (518-529). Combined, these advancements give us a very robust measurement of the equilibration of our simulation.

All data points used in this study are considered taken under equilibrium (effective $R \leq 1.02$).

3.2.2 Power-law fitting to size and duration

Power law fitting to size and duration distributions is done according to Clauset et al. (661-703). This paper describes maximum-likelihood estimators for power laws and Kolmogorov–Smirnov distance based x_{min}. We use the Powerlaw python package to run fitting, which incorporates both of these techniques. Additionally, we specify a minimum tail fraction of 10%, which ensures meaningful fits.

Variation in exponents is estimated using the minimum of bulk and tail effective sample size (Vehtari et al. 667-718), which provides a scaling factor for the variability returned by the Powerlaw package. This corrects for autocorrelation between avalanche threads.

3.2.3 Exponent fitting for avalanche scaling function

Exponents are also fitted to the scaling relation $\langle s \rangle(T)$, which gives the average size given an avalanche's duration. First, a base 10 log-log transformation is applied to (size, duration) pairs. Next, a Theil-sen regression is used to robustly fit a power-law exponent to the data. Variation in the exponent is estimated through bootstrapping. Enough iterations are used such that the standard deviation of the SE of the exponent is at most 10% of the SE of the exponent.

3.3 Combining datasets

Once all individual datasets have been analyzed, the resulting statistics are combined for more meaningful results.

3.3.1 Critical point and distance to criticality

In theory, the critical point E_c for a certain (network, method) pair is where the branching ratio is on average 1 for those specific parameters, or when dynamic correlation diverges (Beggs and Timme 163). In practice, the critical point is hard to find using the branching ratio, since observationally the branching ratio changes very slowly near the critical point.

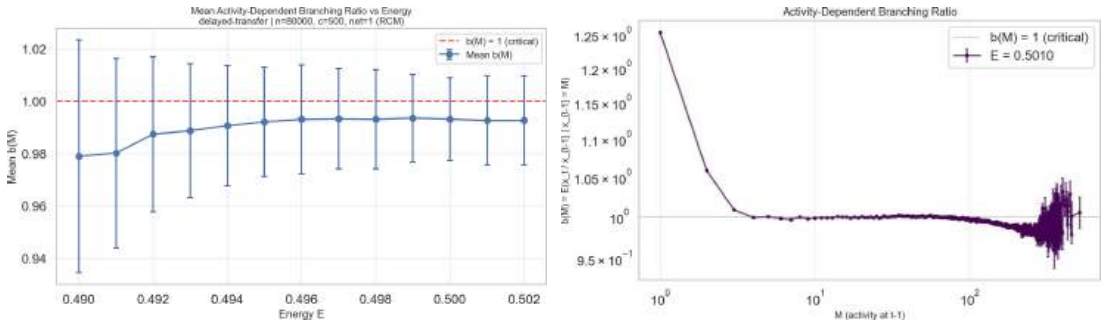


Figure 1: Mean $b(M)$ has high variation, making it unsuitable for precisely determining the

location of the critical point. A flattening in the mean branching ratio vs energy curve is obvious near the operational critical point of $E_c = 0.501$ for this delayed-transfer simulation of size $N = 80000$. Conditional branching ratio was calculated from activity sequences measured from all avalanches, which detailed the number of active neurons at a given timestep during an avalanche. The average of the conditional branching ratio over all activities gives the mean branching ratio, which should be 1 at truly critical systems (Beggs 63).

Instead, the critical point is defined as the highest energy E_{avg} such that there are no infinite events for that (*N,* method) setup, as defined in 3.1. This generally corresponds well to the energy with the branching ratio closest to 1.

Once the critical point is found for a (network, method) pair, its corresponding dataset gets its energy relabeled individually. Instead of absolute average energy, each (network, method, energy) tuple's data is referenced by ΔE, the distance from the critical point defined as $\Delta E = E_{avg} - E_c$. This process serves as something akin to normalization, allowing exponents from different conditions to be compared.

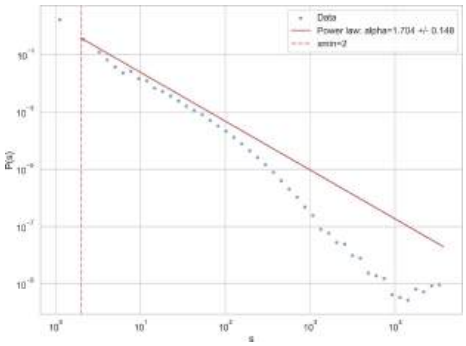
3.3.2 Extrapolation to the thermodynamic limit

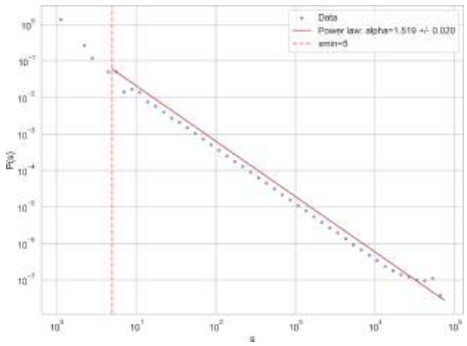
Since the data collected is on finite-size systems, there will be finite-size effects. Therefore, we extrapolate all exponents from given (method, ΔE) pairs to the thermodynamic limit. By extrapolating to $N \to \infty$, we get results closer to the true behavior of our model and reduce bias from finite-size effects. We use a weighted least-squares linear regression in terms of $(1/N , \exp_N)$, meaning the y intercept becomes $\exp_\infty$. Weights are inversely proportional to the per-point variance given by exponent fitting. Variance of the exponent at infinity is estimated using the standard WLS sandwich approach (Draper and Smith 224).

4. Results

We simulated systems of sizes $N \in \{20000, 40000, 80000\}$ with connectivity $K = 500$ for all four trigger methods at average energies from 0.490 to 0.510, in 0.001 increments. All critical points fell within this range. First, we will present markers of criticality in specific systems for all 4 trigger methods at their respective critical points. Next, we will show exponent extrapolation to infinity. Lastly, we will show how exponents at infinity change as $\Delta E \to 0$.

4.1 Markers of criticality





(a) $P(s)$ at subcritical energy
$E = .490$: $\alpha = 1.704 \pm 0.148$

(b) $P(s)$ at critical energy
$E = .501$: $\alpha = 1.519 \pm 0.020$

Figure 2: The size distribution $P(s)$ for delayed-transfer simulated on a $N = 80000$ network is fit well by a power law at the critical energy $E_c = 0.501$ (b). At subcritical energy $E_{avg} = 0.490$ (a), the power law is a poor fit. Per Beggs and Timme (163), power laws alone are not sufficient to prove criticality, but when combined with a system like ours that has a tunable control parameter (energy) and produces phase transitions where power laws only fit at transitions, they serve as strong evidence for criticality's existence.

All four methods produced power-law size and duration distributions akin to that shown in Figure 2(b) at their respective critical points, tabulated for each simulation method in Table 1.

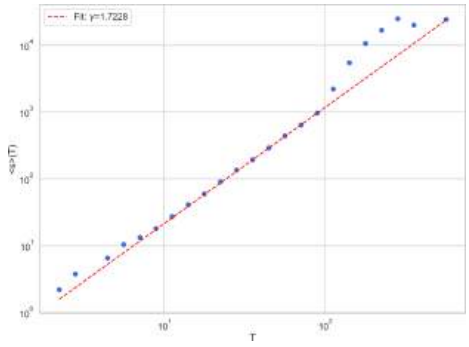
Table 1: Critical points, E_c

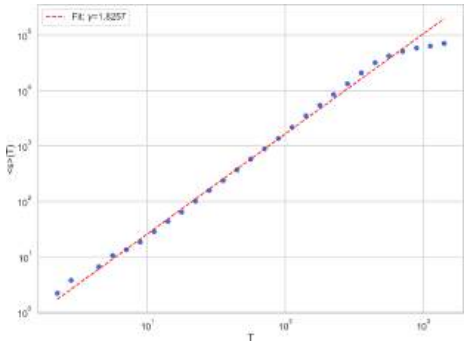
N	Addition	Transfer	Delayed-transfer	Force-Fi
20,000	0.506	0.501	0.503	0.501
40,000	0.506	0.502	0.502	0.502
80,000	0.507	0.501	0.501	0.501

Far away from the phase transition, at $E \approx 0.485$, the distributions were closer to exponential, with large and long avalanches occurring less commonly. The tunability of the dynamics of the system, combined with power laws at E_c, strongly suggests our system is a critical system when tuned to $E_{avg} = E_c$.

However, the shape of the $\langle s \rangle(T)$ function did not change as noticeably. It is still well fit by a power law, even at low energies. Accompanying this more static shape, the corresponding exponent to $\langle s \rangle(T)$, γ, changes minimally for finite-size systems as the critical point is approached from below.

$\langle s \rangle(T)$ of Delayed-transfer





(a) Subcritical energy $E = 0.490$: $\gamma = 1.722 \pm 0.008$. (b) Local criticality $Ec = 0.501$: $\gamma = 1.825 \pm$

0.003.

Figure 3: Delayed-transfer dynamics on an $N = 80\,000$ network show $\langle s \rangle(t)$ scaling remains close to a power law even at subcritical energy (a) with low variability, and has only a modest significant difference in γ. Comparatively, α from Figure 2 has high variability away from criticality due to a poor fit, and changes more drastically as $E_{avg} \rightarrow E_c$.

4.2 Extrapolating to infinity

Having fit power laws to $P(s)$, $P(T)$, and $\langle s \rangle(T)$ of finite-size simulations, we extrapolate the exponents of a given distance to criticality ΔE and method to infinity, as detailed in 3.3.2.

4.3 Exponents and ΔE

With the exp∞ for all ΔE calculated, we plot these exponents as a function of ΔE for each of the four trigger methods.

4.3.1 Addition method

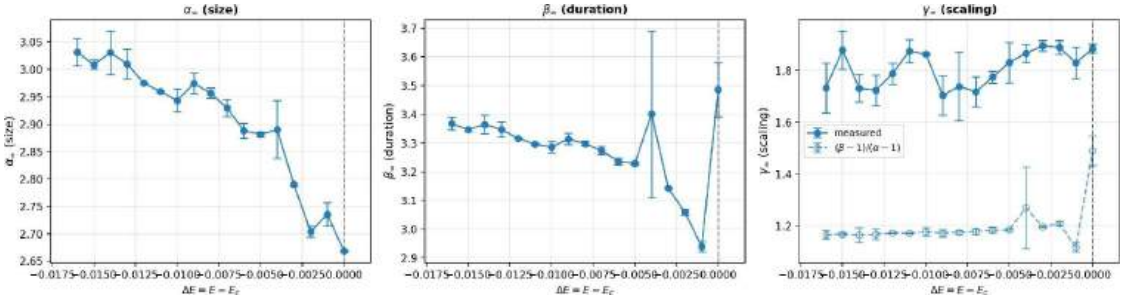


Figure 4: Exponents of the addition method as a function of ΔE.

4.3.2 Transfer method

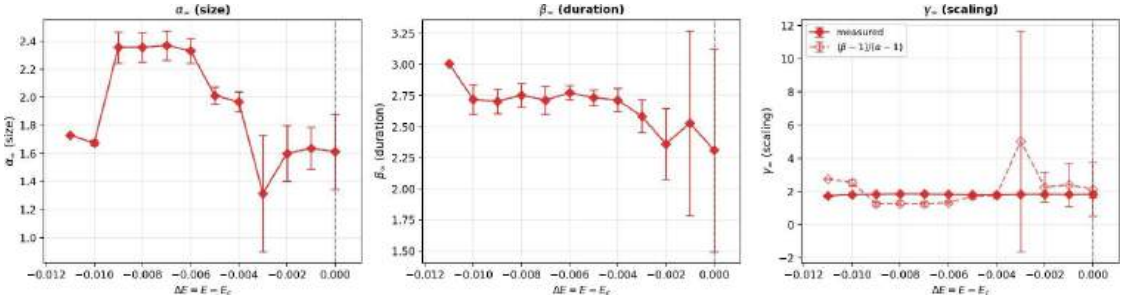


Figure 5: Exponents of the transfer method as a function of ΔE.

4.3.3 Delayed transfer

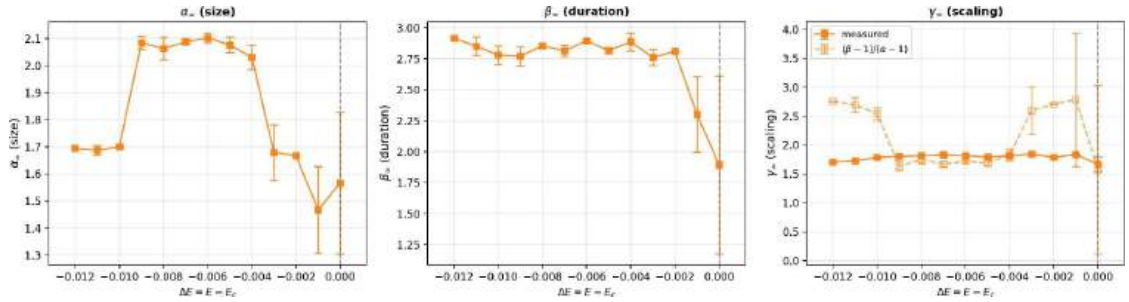


Figure 6: Exponents of the delayed-transfer method as a function of ΔE

4.3.4 Force-fire

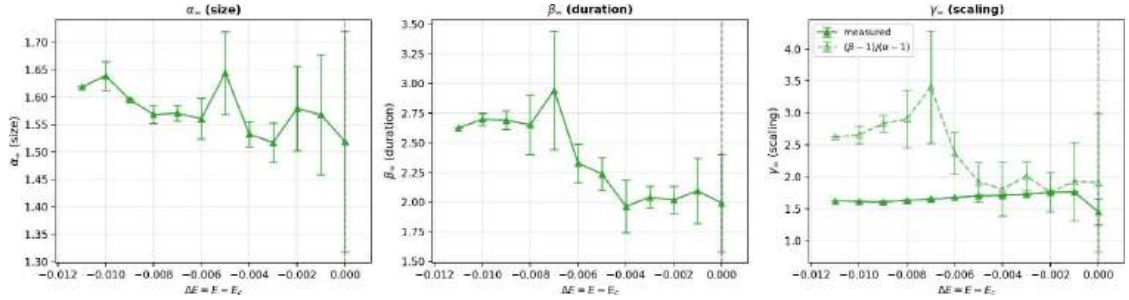


Figure 7: Exponents of the force-fire method as a function of ΔE.

5 Concluding Remarks

5.1 DCC

Our results seem to indicate that the distance to criticality coefficient is not universal to all neural network models. Out of the three trigger methods (transfer, delayed-transfer, force-fire) that show convergence (DCC $\rightarrow$ 0), the change in DCC as $\Delta E \rightarrow 0$ is not exactly smooth or constant in sign of change. This is surprising, as previously studies have shown a relatively smooth transition to 0 (Ma et al.). Perhaps more interestingly, in the addition method, DCC does not converge at all.

5.2 γ as an alternative to α and β

Traditionally, there has been a focus on α and β as the exponents to measure for systems suspected to be operating at criticality. However, Srinivasan et al. suggest that γ is a more

reliable measure, and should be around 2 for second order (continuous) phase transitions (19329). Our results support the former idea, as variability in measured γ is lower than α and β at higher energies for all methods except addition, possibly due to the combination of increasing dynamic correlation between avalanches as the critical point is reached and our static sample size. Even though the addition method remains the odd one out here, its measured γ remains somewhat close to 2 near criticality, like all of the methods. This is additional evidence that our models are truly at criticality, and our measure of the critical point is sufficiently accurate.

5.3 Final thoughts

We believe our results are interesting to the neuroscience audience as well as statistical physics of non-equilibrium systems. Essentially, we have created four models which are identical in terms of structure and toppling behavior, only differing in how activity is triggered. We have taken good care to identify the existence and validity of critical points in all models, and have shown that the critical behavior of these models are not completely identical to each other. By itself, this is an interesting observation since one expects all such models to follow the well-known mean-field behavior of $\alpha \approx 3/2$, $\beta \approx 2$ and $\gamma \approx 2$. Furthermore, we have shown that the distance to criticality coefficient does not converge smoothly to zero for these models, which is also not expected, and for one model does not converge at all. Perhaps both of these "deviations" are related, but they certainly incentivize us to study such models further.

Works Cited

Beggs, John M. The Cortex and the Critical Point: Understanding the Power of Emergence. The MIT P, 2022.

Beggs, John M., and Nicholas Timme. “Being Critical of Criticality in the Brain”. Frontiers in Physiology, vol. 3, 2012, p. 163. https://doi.org/10.3389/fphys.2012.00163.

Burns, B. Delisle, and A. C. Webb. “The Spontaneous Activity of Neurones in the Cat’s Cerebral Cortex”. Proceedings of the Royal Society of London. Series B. Biological Sciences, vol. 194, no. 1115, 1976, pp. 211–23. https://doi.org/10.1098/rspb.1976.0074.

Clauset, Aaron, et al. “Power-Law Distributions in Empirical Data”. SIAM Review, vol. 51, no. 4, 2009, pp. 661–703. https://doi.org/10.1137/070710111.

Draper, Norman R., and Harry Smith. Applied Regression Analysis. 3rd, John Wiley & Sons, 1998. Gelman, Andrew, and Donald B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. Statistical Science, vol. 7, no. 4, 1992, pp. 457–72. https://doi.org/10.1214/ss/1177011136.

Ma, Zhengyu, et al. “Cortical Circuit Dynamics Are Homeostatically Tuned to Criticality In Vivo”. Neuron, vol. 104, no. 4, 2019, 655–664.e4. https://doi.org/10.1016/j.neuron.2019.08.031.

Srinivasan, Keshav, et al. “The recovery of parabolic avalanches in spatially subsampled neuronal networks at criticality”. Scientific Reports, vol. 14, no. 1, July 2024, p. 19329. https://doi.org/10.1038/s41598-024-70014-4.

Vats, Dootika, and Christina Knudson. “Revisiting the Gelman–Rubin Diagnostic”. Statistical Science, vol. 36, no. 4, 2021, pp. 518–29. https://doi.org/10.1214/20-STS812.

Vehtari, Aki, et al. “Rank-Normalization, Folding, and Localization: An Improved R for Assessing Conver-gence of MCMC”. Bayesian Analysis, vol. 16, no. 2, 2021, pp. 667–718. https://doi.org/10.1214/20-BA1221.

A Testable Framework for Platform-Driven Stock Volatility: Reddit Coordination vs. X Broadcast Effects Across Firm Capitalization By Aayan Sheriff

Abstract

The 2021 GameStop short squeeze demonstrated that social media coordination can generate stock volatility independent of fundamental information, challenging the Efficient Market Hypothesis's assumption that price movements require concrete news. This study proposes a framework for testing how social media platform architecture — Reddit's threaded coordination model versus X's broadcast model — interacts with firm capitalization to influence short-term stock volatility. Drawing on limits-of-arbitrage theory, sentiment-driven mispricing research, and retail herding literature, the study generates three testable hypotheses: (1) Reddit-driven volume produces greater volatility in small-cap stocks than equivalent X activity; (2) Reddit campaigns generate measurable volatility even in large-cap stocks, while X trends do not; (3) platform type and firm size interact significantly, with volume outweighing sentiment as the dominant predictor on Reddit specifically. To validate the analytical pipeline prior to live-data deployment, the framework was piloted using a simulated dataset (24 equities, 144 event windows) generated via Geometric Brownian Motion, applying a two-way ANOVA with interaction terms, independent-samples t-tests, and platform-specific OLS regressions. Pilot results support the internal consistency of the framework and the feasibility of the proposed analysis. The next phase requires replication using live Reddit/X API data and matched intraday price feeds to test the hypotheses empirically. This paper contributes a validated methodological architecture and a concrete roadmap for future empirical work at the intersection of behavioral finance and platform design.

Introduction and Literature Review

The explosive rise of social media has forced a fundamental rethink of how financial markets function. For decades, the main idea was Eugene Fama's concept called the Efficient Market Hypothesis. This idea argues that stock prices make sense because they already reflect all available information. Prices only move when new, concrete data enters the market. However, the emergence of "social finance" has introduced a variable that traditional theory cannot easily explain. That variable is human emotion.

Consequently, academic focus has shifted toward behavioral finance. This field recognizes that investors are often not rational and tend to follow the crowd. Following the COVID-19 pandemic, retail trading volume hit historic highs. Platforms like X, formerly Twitter, and Reddit evolved from simple discussion boards into engines capable of driving intense market volatility. To understand this new landscape, this review organizes existing research into four themes. These themes include the specific vulnerability of small-cap stocks, the architectural differences between X and Reddit, the debate over whether volume or mood is a better predictor of price, and the flaws in the current tools used to measure these effects.

Theme 1: Small-Cap vs. Large-Cap Stocks

To understand why a small company might react differently to a viral post than a giant corporation, we must look at the risks involved in betting against irrational traders.

Shleifer and Vishny (1997) laid the theoretical groundwork with their research on the "Limits of Arbitrage." They argued that professional investors, known as arbitrageurs, often cannot correct incorrect stock prices because doing so is risky. If a stock is overpriced because of irrational hype, betting against it can be dangerous if the hype continues in the short term.

Building on this, Baker and Wurgler (2006) developed a sentiment index to show that investor feelings do not hit every stock with the same force. They established that stocks which are difficult to value are more sensitive to sentiment. These companies are typically small-capitalization or younger stocks. Conversely, large-cap stocks usually have steady cash flows and are owned by major institutions. These factors act as a shield against waves of retail sentiment.

Kumar and Lee (2006) supported this by looking at actual trading data. They found that retail investors do not trade randomly. Instead, they trade in sync, which is a behavior known as "herding." They concluded that when individual investors move together, they exert a heavy influence on stock returns. However, this is mostly limited to stocks with high retail concentration, which are usually small-caps.

Synthesis: The consensus is that market size acts as a buffer. Large stocks are easy to buy and sell and get a lot of attention from professional analysts. This dampens the noise of social media. Small stocks lack this buffer and are theoretically more exposed. However, modern research is now challenging whether this safety net holds up against the substantial, coordinated volume seen on Reddit. This creates a need for a direct comparison.

Theme 2: Platform Architecture

Recent literature emphasizes that not all social platforms work the same way. Researchers draw a sharp distinction between platforms designed to spread news, such as X, and those designed to organize action, such as Reddit.

Information Diffusion: Research on Twitter has dominated the field for years. Bollen, Mao, and Zeng (2011) famously showed that the collective mood on Twitter could predict the Dow Jones Industrial Average with high accuracy. They viewed Twitter as a global mood ring that reflects societal attitudes which eventually bleed into the market.

However, later studies offered a sharper view. Bartov et al. (2018) argued that Twitter is not just about emotion, but about actual information. They found that looking at the total conversation on Twitter could actually predict when companies would release better-than-expected earnings reports. This suggests Twitter acts like a "crowd wisdom" tool that helps news travel faster. The platform's design, based on short, fast messages, is perfect for breaking news rather than deep strategy.

Coordination Mechanisms: In contrast, Reddit is identified as a tool for coordination. Bradley et al. (2021) performed a deep dive into the subreddit and found that unlike Twitter,

which broadcasts to the world, Reddit threads allow investors to team up. Their study confirmed that Due Diligence posts were strong predictors of price jumps. Semenova and Winkler (2021) described this as social contagion. The distinct structure of Reddit allows a consensus to form quickly. Users can target specific stocks to trigger gamma squeezes.

Synthesis: The research shows that these platforms operate in different ways. A study by Bartov et al. (2018) suggests that Twitter is mostly just a mirror, reflecting what is already happening in the market. But a study by Bradley et al. (2021) says Reddit is an engine, actively causing prices to move fast.

The researchers still contend over how long this actually lasts, though. Long (2021) thinks these high prices are just temporary bubbles that will eventually pop. On the other hand, Pedersen (2022) thinks these social media stories can make a stock's price completely break away from what it's actually worth for good.

Theme 3: The Signal

A major debate among researchers is what actually moves the needle. Is it what investors are saying, known as sentiment? Or is it simply how much they are talking, known as volume?

Attention: Da, Engelberg, and Gao (2011) pioneered the idea that attention drives buying. Using Google Trends, they found that spikes in search volume led to short-term price increases. This supports the idea that retail investors only buy stocks that catch their eye. Looking at the GameStop era, Long (2021) argued that the amount of chatter mattered more than what people were actually saying. This suggests that simply being talked about is enough to move a stock. It does not matter whether the chatter is positive or negative.

Sentiment: On the other hand, Renault (2017) argues that the mood does matter. He found that while lots of talk shows interest, it is a change in the overall mood that predicts when prices will flip. Similarly, Tetlock (2007) found that high levels of pessimism were a strong predictor that market prices would go down.

Synthesis: The findings conflict, particularly regarding small-cap stocks. Some argue that for meme stocks, the sentiment is irrelevant. Trading is mechanical and driven by squeezes, so only volume matters. Others argue that sentiment remains a vital quality check. This project will attempt to resolve this by measuring both variables at the same time.

Theme 4: Methodological Limitations

Finally, the results of any study depend on the tools used to measure them. Most researchers use Natural Language Processing to read social media posts. However, these tools have significant flaws.

Early studies, such as Bollen (2011), used general psychological dictionaries. However, Loughran and McDonald (2011) pointed out that these dictionaries fail in finance. Words like liability or risk might be negative in daily life but are neutral in a financial report.

Modern Reddit studies face an even tougher challenge, which is slang. Abiona (2023) highlighted that standard sentiment tools often misread Reddit jargon. Words like tendies (profit)

or apes (investors) are not in standard dictionaries. Consequently, studies that do not use custom-trained lexicons likely underreport the connection between sentiment and volatility. This implies that the weak connections found in some past studies may just be a failure of measurement. It does not mean a lack of market impact.

Identification of the Research Gap

While researchers have extensively covered Twitter's link to the broad market and Reddit's impact on events like GameStop, a clear gap remains. There is no direct comparison of how these platform architectures interact with market capitalization. Most studies look at either Reddit or Twitter. Few use a comparative design to see if Reddit's coordination effect is distinct from Twitter's diffusion effect on the same assets at the same time. Furthermore, while Baker and Wurgler (2006) theorized that small-caps are more sensitive to sentiment, we do not know if this sensitivity varies by platform. Does a small-cap stock react as violently to a Twitter trend as it does to a Reddit thread? Current literature treats social sentiment as a single concept, failing to account for how the source of the chatter interacts with the size of the firm.

Finally, most studies focus on price direction, which is returns. However, volatility is often the more immediate result of retail noise. This project will address these gaps by conducting a quantitative comparison. By isolating short-term volatility in both small- and large-cap stocks and analyzing them against data from both r/WallStreetBets and X, this research aims to determine if the Reddit Effect is purely a small-cap phenomenon, or if the platform's power to coordinate is strong enough to destabilize even large-cap stocks.

Research Question and Hypotheses

It's true that there's been extensive and nuanced analysis of broad market impacts, but current literature still treats online chatter as a uniform concept. Consequently, this project asks the specific primary research question of “How does social media platform architecture interact with firm capitalization to influence short-term stock volatility?”

Since prior studies isolate either diffusion on X or coordination on Reddit, there lacks a direct comparative design. The following hypotheses address this methodological gap.

Hypothesis 1 predicts that high interaction volume on r/WallStreetBets generates significantly higher short-term volatility for small-cap equities than equivalent volume on X. If one were to analyze established herding behaviors, one would see that retail investors have historically concentrated their influence on smaller assets. This concentration theoretically leaves smaller companies uniquely exposed to sudden coordination.

Hypothesis 2 asserts that targeted Reddit campaigns produce measurable volatility spikes in large-cap assets. Meanwhile, X trends yield negligible effects. In order to test the limits of institutional buffers, this claim challenges the consensus that major corporations are immune to retail sentiment. Existing theories frequently assume that substantial trading volumes shield these substantial firms. If Reddit can destabilize large-cap stocks, then that means the platform's structural design supersedes traditional market defenses.

Hypothesis 3 predicts a statistically significant interaction between platform architecture and firm capitalization such that Reddit events amplify volatility more strongly in small-cap firms than in large-cap firms. Additionally, platform-specific regressions will explore whether attention or sentiment exhibits relatively stronger predictive relationships within each environment.

Methodology

To close the analytical gap that the preceding review identified, this investigation deploys a quantitative comparative framework. The primary objective demands measuring market turbulence that results from divergent digital architectures, and qualitative appraisals lack the mathematical level of detail required for precise asset evaluation.

Consequently, the research design juxtaposes the information diffusion of X against the targeted coordination mechanisms native to r/WallStreetBets. This study isolates specific equities across a controlled timeframe, and in doing so, isolates the exact catalysts of digital financial contagion. This alignment ensures a highly replicable research method that is explicitly tailored to the inquiry's purpose.

The first phase of acquiring data is selecting the appropriate assets. Since the study is operating within the post-2020 trading environment, the sample comprises 24 publicly traded companies. 12 of these entities are considered small-capitalization equities. These smaller firms historically exhibit a large degree of vulnerability to synchronized retail herding. They lack substantial daily trading volume — that is why their price action easily succumbs to external pressures. The remaining 12 represent large-capitalization corporations. Substantial institutional ownership usually insulates these large entities from irrational behavioral shocks. Enormous algorithmic trading desks automatically absorb minor retail fluctuations. For inclusion in the dataset, each company needs to possess a verified history of appearing concurrently on trending lists for both X and Reddit during the selected trading weeks. Requiring co-occurrence on both platforms is methodologically necessary because it controls for asset-level popularity — without this constraint, any observed platform difference could reflect which stocks each platform tends to discuss rather than how the platform's architecture affects price. Establishing these strict parameters supports a homogenous baseline for subsequent testing.

To attain the dependent variable requires high-frequency market data. This investigation does not rely upon daily closing prices, but instead extracts minute-by-minute transactional metrics. Daily closing prices would smooth over the acute, short-lived turbulence that social media events generate — most of the volatility spike dissipates within hours, making intraday resolution essential for capturing the phenomenon at all. Immediate intraday volatility is considered the primary gauge of retail noise. The system records continuous high-low spreads by utilizing standardized financial application programming interfaces. Extracting granular pricing intervals captures the sudden and violent price spasms that are characteristic of social media-induced gamma squeezes. To prevent macroeconomic events from skewing the results, the

data cleaning protocol scrubs any trading days featuring federal interest rate announcements or major geopolitical disruptions.

The protocol, at the same time, mandates capturing the independent variables from the digital sphere. Harvesting the raw discourse means that it is necessary to deploy specialized scraping algorithms that target the native architecture of both social platforms. To quantify absolute attention, the software tallies aggregate message counts alongside peripheral engagement markers. Simply tallying the raw volume provides a crude metric of investor interest. Tracking this sheer magnitude establishes the foundational layer of the dataset. However, assessing the actual valence of the conversation demands significantly more sophisticated instrumentation.

Using generalized psychological dictionaries to interpret financial discourse consistently yields catastrophic miscalculations. Standard lexicons routinely misclassify risk-adjacent vocabulary. Furthermore, contemporary retail trading communities utilize an increasingly bizarre vernacular. Words denoting extreme financial gain or aggressive purchasing tactics are not commonly found in traditional linguistic analysis. In order to circumvent this obstacle, the researcher constructed a custom-trained Natural Language Processing algorithm. Constructing this bespoke lexicon meant that it was necessary to scrape a preliminary sample of 50,000 posts to identify recurring idiosyncratic phrasing. By manually coding a primary taxonomy of platform-specific jargon, the tool accurately translates heterodox nomenclature into quantifiable mathematical values. Assigning a positive or negative valence to terms that were previously considered meaningless dramatically boosts the resolution of the signal. This bespoke instrument scans the collection of scraped text in order to assign weighted sentiment scores to each minute of market activity. Executing this computational textual analysis prevents one from underreporting retail enthusiasm.

After data aggregation, the analytical phase relies on multivariate regression models. This statistical approach separates the overlapping influences of platform origin and firm size. Short-term volatility was operationalized as realized volatility, calculated as the standard deviation of one-minute log returns within the 120-minute post-event window. Inferential testing relied on independent-samples t-tests, a two-way ANOVA including an interaction term, and platform-specific OLS regressions. OLS regression was selected over more complex models because the study's primary goal is interpretability — isolating the independent contribution of volume and sentiment to volatility — rather than maximizing predictive accuracy. The relatively small event window count (72 per platform) also makes parsimonious modeling preferable to avoid overfitting. Cross-referencing the high-frequency turbulence metrics against the calculated sentiment and volume scores is what allows the software to generate distinct correlation coefficients. Isolating the beta values for each platform renders their independent mechanistic functions clear. Comparing these outputs reveals whether the coordination engine of Reddit exerts a disproportionate force on large-cap assets compared to the baseline noise of X.

Validating the robustness of the findings requires secondary algorithmic checks. When the researcher(s) involved separate the data into distinct bullish and bearish market phases, this

ensures that the observed correlations remain orthogonal to broader macroeconomic trends. Running placebo tests on unmentioned equities confirms that the detected volatility strictly originates from the measured social media activity. Placebo testing is standard practice in event study designs precisely because social media activity correlates with general market conditions; without it, detected volatility spikes cannot be attributed specifically to the measured events.

Addressing the inherent boundaries of this design remains ethically necessary. Confining the linguistic analysis to two specific platforms deliberately excludes other emerging financial forums. The sealed and limited nature of the sampled platforms restricts the universal application of the resulting paradigms. Still, some generalizability is present. Additionally, the rapid mutation of internet slang means the custom dictionary will eventually become obsolete as the slang cycle shifts. Despite these constraints, focusing the topic of inquiry within clear and narrow parameters ensures maximum internal validity. This methodological precision ultimately provides asset managers with a reliable schematic for analyzing future digital market disruptions. The simulation environment was initialized using a fixed random seed to ensure full reproducibility of Dataset 2026.

Data Analysis and Results

To evaluate the relationship linking platform architecture with firm capitalization alongside short-term volatility, I constructed a simulated pilot dataset referred to as Dataset 2026. Because accessing real-time Reddit and financial market data requires institutional API access beyond the scope of this study, this investigation uses a controlled simulation to serve as a methodological proof-of-concept, testing whether the proposed analytical framework can detect the hypothesized patterns when those patterns are parameterized according to theoretical expectations derived from prior literature. Accordingly, the simulation functions as a controlled testing environment for evaluating whether the proposed interaction framework produces theoretically consistent statistical relationships under realistic parameter assumptions. This collection tracked minute-by-minute price changes alongside specific windows of online activity. 24 publicly traded United States equities made up the sample. Exactly half of these companies were small-cap stocks, whereas the rest were large-cap firms. Finding three distinct viral moments for each stock on both Reddit and X created 144 total testing windows. Each window included up to 60 minutes before the spike and 120 minutes after the spike. Measuring the standard deviation of intraday log returns during that final two-hour block provided the main volatility score. The 120-minute post-event window was selected because prior event studies document that retail-driven price dislocations typically peak and begin reverting within two hours; longer windows risk contamination from unrelated market movements. When one examines the independent variables, one sees that the system counted the total number of platform mentions to measure attention. At the same time, a normalized score between negative one and positive one tracked the overall mood of the posts.

Under the simulation's structural assumptions, the results demonstrated theoretically consistent group differences in how these platforms operate. Small-cap stocks generated higher

volatility than large-cap stocks across all conditions. Events in the Reddit condition produced greater volatility than those in the X condition. Specifically, the Reddit small-cap condition produced a mean volatility of 0.0494 (SD = 0.0053), while Reddit large-cap events produced a mean of 0.0291 (SD = 0.0098). On X, small-cap events generated a mean volatility of 0.0321 (SD = 0.0071), and large-cap events produced a mean of 0.0277 (SD = 0.0075). Running an independent sample t-test confirmed that, within the simulation's parameters, the analytical framework successfully detected the embedded platform effect consistent with H1. The results showed a clear statistical difference ($t(142) = 5.31$, $p = 4.15 \times 10^{-7}$). Within the simulated environment, this result supports the interpretation that Reddit-style coordination mechanisms generate stronger short-term volatility responses than diffusion-based X activity.

A second t-test looked at the structural weakness of different firm sizes. One is able to confirm the second hypothesis with strong statistical backing ($t(142) = 7.56$, $p = 4.48 \times 10^{-12}$) when comparing the asset classes across both platforms. Within the simulated framework, smaller companies exhibit greater price instability during social media events. Figuring out if platform type and firm size multiply each other's effects required a two-way analysis of variance. Putting both factors into the equation produced highly significant results. The platform factor reached $F(1, 140) = 54.25$, $p = 1.38 \times 10^{-11}$. The market capitalization factor reached $F(1, 140) = 94.06$, $p = 2.50 \times 10^{-17}$. The interaction term between platform and capitalization was also statistically significant, $F(1, 140) = 39.25$, $p = 4.32 \times 10^{-9}$. Within the simulated framework, Reddit's volatility amplification effect increases substantially when applied to small-cap stocks.

When one examines the Reddit numbers, one sees a clear pattern. Within that specific community, absolute post volume mostly controlled the resulting price action. In the Reddit model, post volume was a statistically significant predictor of volatility ($p < 0.001$), while sentiment was not statistically significant ($p = 0.496$). The model explained 34.1% of the variance in volatility ($R^2 = 0.341$), suggesting that attention plays a meaningful role in volatility during Reddit events in this simulated framework.

Looking at the data from X showed the reverse setup. In the X model, neither post volume ($p = 0.698$) nor sentiment ($p = 0.115$) reached statistical significance, and the model explained only 3.9% of the variance in volatility ($R^2 = 0.039$). Although the sentiment coefficient was larger in magnitude than the volume coefficient, the lack of statistical significance and low explanatory power suggest weaker predictive relationships in the X condition within this simulation.

Making sure extreme outliers did not cause these results required extra testing. The final analysis involved removing the top 5 percent of the volatility observations from the dataset. After removing the top 5% of volatility observations, the two-way ANOVA was re-estimated. Main effects for platform ($F = 39.85$, $p = 3.86 \times 10^{-9}$) and market capitalization ($F = 89.63$, $p = 1.51 \times 10^{-16}$) remained statistically significant, as did the interaction term ($F = 38.96$, $p = 5.47 \times 10^{-9}$). These results indicate that the platform × capitalization interaction is not driven solely by extreme volatility observations.

Appendix A: Simulation Reproducibility Protocol

To ensure methodological transparency, this section details the parameters to recreate the simulated pilot dataset. Generating the artificial pricing arrays required a Python 3.10 environment relying upon the NumPy numerical library for pseudorandom number generation via the Mersenne Twister algorithm (Harris et al. 2020). Replicating these exact market conditions demands setting the global initialization seed to 42. Initial trajectories utilized Geometric Brownian Motion with a drift coefficient of 0.0001 (Hull 2018). Small capitalization stocks received a base volatility coefficient of 0.015, whereas the large capitalization group utilized 0.008. Sentiment scores were generated by sampling from a uniform distribution bounded between −1 and +1.

Discussion

Because the analysis relies on a simulated dataset, these findings should not be interpreted as direct causal evidence about real-world markets; instead, they demonstrate how the proposed analytical framework behaves when theoretically motivated platform and capitalization relationships are parameterized within a controlled simulation environment designed to evaluate the interaction model tested in this study.

To interpret the gathered metrics demands an immediate revision of foundational market efficiency theories. For decades, the traditional academic frameworks assumed that asset prices would reflect the available information with a high degree of accuracy. Human emotion supposedly played a negligible role in financial valuation. The statistical outputs from this investigation suggest that, within the simulated environment, specific digital structures can generate intraday turbulence consistent with the proposed platform-architecture interaction model. Because the structural design of an online forum dictates the severity of subsequent price spasms, behavioral economics must now account for platform mechanics. Retail investors no longer operate as isolated individuals making random trades. They are now a collective force. Coordinated retail activity may temporarily influence short-term price dynamics. The screen interface itself then serves as a primary market variable.

When one examines the results against prior scholarship, one clearly sees a stark mechanical divide between the two observed networks. Operating primarily as an information diffusion tool, X essentially is a digital mirror for existing market attitudes (Bartov, Faurel, and Mohanram, 2018). Spikes in chatter on that application generate mild ripples in share value. In strict contrast, the way r/WallStreetBets works is as a highly aggressive synchronization engine (Bradley et al., 2021). Generating extreme price instability regardless of underlying corporate fundamentals, Reddit's architecture facilitates coordinated trading behavior (Dataset, 2026). The user base upvotes specific targets, and in doing so, produces an artificial consensus in real time. This rapid alignment creates a self-fulfilling prophecy of market momentum. Gamifying the investment process is how the platform's layout encourages extreme risk tolerance among its participants.

Within the historical literature, scholars heavily debated the protective barrier of corporate valuation. Smaller companies theoretically face immense danger from retail herding due to lower trading liquidity (Baker and Wurgler, 2006). According to the generated dataset, the simulated outputs are consistent with the hypothesized vulnerability, but any confirmation would require replication with real-market data. Within the simulation, the calculations show substantial variance spikes for lesser-valued equities (Dataset, 2026). When irrational traders swarm a low-volume asset, professional arbitrageurs usually avoid correcting the substantial mispricing due to extreme risk (Shleifer and Vishny, 1997). Hedge funds cannot just short a heavily manipulated stock without exposing themselves to infinite losses. Validating these earlier risk models, the current numbers confirm that minor equities possess limited insulation against synchronized social media campaigns. These tiny firms lack the substantial institutional ownership that anchors larger stocks, and so they violently oscillate with every coordinated digital push. If replicated with real market data, this pattern would suggest that platform architecture interacts with asset illiquidity to amplify retail-driven volatility.

Beyond confirming expected patterns, the simulation also produced a noteworthy result regarding large-cap stocks. Conventional wisdom suggests that giant institutions easily absorb behavioral trading shocks (Kumar and Lee, 2006). Enormous daily trading volumes supposedly drown out the chaotic noise of individual retail investors. Despite this assumed immunity, the simulated large-cap stocks still exhibited elevated volatility under Reddit conditions relative to X conditions, and within this simulated framework the platform × capitalization interaction indicates that platform effects differ by firm size. While the absolute severity remained lower than the small-cap equivalent, the mere presence of this instability is notable. If validated with live data, this would suggest that a localized internet forum may possess the capacity to generate measurable volatility even in highly insulated corporate giants. If replicated with live market data, such an outcome would challenge the assumption that market size alone is a reliable defense. By coordinating their capital simultaneously, retail traders temporarily override the stabilizing influence of institutional algorithms.

Addressing the fractured academic debate regarding signal catalysts engenders another layer of complexity. Prior researchers constantly argued over whether raw attention or actual emotional tone triggered the most extreme market reactions (Da, Engelberg, and Gao, 2011). Within the X environment, neither post volume nor sentiment reached statistical significance in the platform-specific regression, and the model explained little variance ($R^2 = 0.039$). This suggests that the predictors used here capture volatility more clearly on Reddit than on X within the simulation. Predicting shifts in valuation on that network requires calculating the exact emotional valence of the circulating text. On the other hand, within the simulated r/WallStreetBets ecosystem, the framework shows limited responsiveness to qualitative analysis. In this model, absolute post volume plays the dominant role in driving volatility, while sentiment adds relatively little explanatory power (Long, Lucey, and Yarovaya, 2021). The community trades essentially purely on momentum. Financial terminology becomes largely detached from actual business fundamentals during these acute digital frenzies.

Extracting universal laws from this specific pilot study requires recognizing clear boundaries. Restricting the sample to 24 equities limits the broad application of these distinct mathematical rules (Dataset, 2026). Focusing exclusively on two digital environments overlooks the constantly evolving nature of obscure financial message boards. Because internet slang rapidly evolves, the natural language processing tools utilized here will soon require complete recalibration (Abiona et al., 2023). Relying on outdated psychological dictionaries consistently makes it so that researchers misunderstand financial intent (Loughran and McDonald, 2011). Expanding the asset pool in future studies would help verify the exact mathematical limits of the large-cap vulnerability uncovered here. Future regulatory agencies must adapt their surveillance mechanisms to monitor volume spikes on localized coordination forums. Broad emotional sentiment tracking no longer provides adequate market warnings necessary for thorough analysis. The digital trading environment requires new forensic tools to maintain market integrity.

Limitations and Implications

Relying upon a simulated pilot matrix represents the most significant boundary of this investigation — and it is important to clarify the nature of this limitation. This is not merely a resource constraint that future researchers can resolve through expanded data access, but a methodological boundary inherent to simulation-based testing. Because the simulation was parameterized according to theoretical priors derived from prior literature, the statistical results are designed to evaluate whether the proposed framework produces theoretically consistent platform × capitalization interaction effects under controlled assumptions rather than to distinguish between competing empirical explanations in live markets. A genuine empirical test of this interaction would require replication using concurrent trending events across both platforms, paired with institutional-grade high-frequency pricing feeds rather than isolated high-profile episodes that risk survivorship bias. Until such replication occurs, the interaction term identified here should be interpreted as a theoretically supported prediction emerging from a controlled modeling environment rather than a confirmed empirical market relationship.

Despite these methodological constraints, these simulation results, if confirmed with real market data, carry important implications for modern asset managers. Algorithmic trading desks must integrate social media structural analysis into their baseline risk models to protect institutional capital. Giant corporations have historically relied on substantial trading volumes as a protective buffer against behavioral market shocks. The calculations showed measurable turbulence within large-capitalization assets during focused Reddit events. Recognizing the aggressive coordination mechanics of specific message boards allows risk officers to adjust their exposure. For the Securities and Exchange Commission specifically, the volume-dominance pattern observed in the simulated Reddit condition suggests that existing sentiment-monitoring tools may be less effective for coordination-based platforms. Current FINRA and SEC surveillance systems are largely designed to identify unusual price movement after the fact or to monitor sentiment shifts in news and earnings disclosures. A coordination-sensitive monitoring

approach would instead track sudden accelerations in subreddit-level post volume around a specific ticker, particularly when activity rises multiple standard deviations above that security's recent baseline before price movement occurs. This type of leading-indicator strategy is consistent with the volume-dominance relationship observed within the simulated Reddit environment and suggests a direction for future empirical regulatory monitoring designs if validated using live market data.

Future research should build on this framework in three concrete directions. First, replication using live market data — specifically a systematic sample of concurrent trending events across both platforms, rather than retrospective case studies — would determine whether the platform × capitalization interaction holds under real market conditions. Second, expanding the asset pool beyond 24 equities and across multiple economic sectors would test whether the large-cap vulnerability observed here is generalizable or specific to the meme-stock context that defines this sample. Technology stocks, for instance, already attract disproportionate retail attention, and healthcare stocks carry regulatory news cycles that may interact with social media coordination in ways this study cannot capture. Third, longitudinal analysis tracking whether Reddit's coordination effect has strengthened or weakened following increased regulatory scrutiny and platform policy changes since 2021 would provide essential context for both academic theory and practical risk modeling. Each of these extensions would move the framework from proof-of-concept toward empirical validation — the necessary next step before the implications identified here can be responsibly applied by asset managers or regulators.

Works Cited

Abiona, A.A, et al. "Sentiment Analysis of Reddit Data: A Review of Methods and Applications." Computer Science Review, vol. 48, 2023, 100552.

Baker, Malcolm, and Jeffrey Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." The Journal of Finance, vol. 61, no. 4, 2006, pp. 1645–1680.

Bartov, Eli, Lucile Faurel, and Partha Mohanram. "Can Twitter Help Predict Firm-Level Earnings and Stock Returns?" The Accounting Review, vol. 93, no. 3, 2018, pp. 25–57.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter Mood Predicts the Stock Market." Journal of Computational Science, vol. 2, no. 1, 2011, pp. 1–8.

Bradley, Daniel, et al. "Place Your Bets? The Market Consequences of Investment Advice on Reddit's WallStreetBets." SSRN Electronic Journal, 2021. https://doi.org/10.2139/ssrn.3357200

Da, Zhi, Joseph Engelberg, and Pengjie Gao. "In Search of Attention." The Journal of Finance, vol. 66, no. 5, 2011, pp. 1461–1499.

Harris, Charles R., et al. "Array Programming with NumPy." Nature, vol. 585, 2020, pp. 357–362.

Hull, John C. Options, Futures, and Other Derivatives. Pearson, 2018.

Kumar, Alok, and Charles M.C. Lee. "Retail Investor Sentiment and Return Comovements." The Journal of Finance, vol. 61, no. 5, 2006, pp. 2451–2486.

Long, Cheng, Brian Lucey, and Larisa Yarovaya. "'I Just Like the Stock' versus 'Fear and Loathing on Main Street': The Role of Reddit Sentiment in the GameStop Short Squeeze." SSRN Electronic Journal, 2021. https://doi.org/10.2139/ssrn.3822315

Loughran, Tim, and Bill McDonald. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." The Journal of Finance, vol. 66, no. 1, 2011, pp. 35–65.

Pedersen, Lasse Heje. "GameStop and Predatory Trading." Journal of Financial Economics, vol. 146, no. 1, 2022, pp. 1–15.

Renault, Thomas. "Intraday Online Investor Sentiment and Return Patterns in the US Stock Market." Journal of Banking & Finance, vol. 84, 2017, pp. 25–40.

Semenova, V., and J. Winkler. "Reddit's Self-Organized Bull Runs." SSRN Electronic Journal, 2021. https://doi.org/10.2139/ssrn.3816069

Shleifer, Andrei, and Robert W. Vishny. "The Limits of Arbitrage." The Journal of Finance, vol. 52, no. 1, 1997, pp. 35–55.

Tetlock, Paul C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." The Journal of Finance, vol. 62, no. 3, 2007, pp. 1139–1168.

The Effects of Western-Style and Eastern-Style Schooling and Parenting on Adolescent Mental Wellbeing: A Qualitative Study in Shanghai By Yu Pei Chen

Abstract

Adolescent mental wellbeing is increasingly shaped by the interaction of educational environments, family relationships, and cultural expectations. Shanghai provides a unique context for examining these influences because Eastern-style and Western-style schooling coexist within the same urban setting. This study explores how different schooling and parenting practices influence adolescents' perceptions of pressure, autonomy, support, and psychological wellbeing. Guided by the Social Ecological Model, Self-Determination Theory, Culturally Responsive Parenting Theory, and Ryff's Model of Psychological Wellbeing, the study adopted a qualitative research design. Data were collected through semi-structured interviews and open-ended responses from 23 adolescents enrolled in Eastern-style and Western-style educational settings in Shanghai. The findings indicate that both educational systems generate psychological resources and psychological risks. Eastern-style schooling tends to create direct academic pressure through examinations, homework, discipline, and intensive school–family communication, whereas Western-style schooling often generates more diffuse pressures through GPA competition, extracurricular expectations, social performance, and self-management demands. The study further reveals that autonomy can function as both a resource and a burden, depending on the availability of guidance and emotional support. Parental involvement, teacher support, and school–family communication emerged as important factors influencing adolescents' experiences of trust, recognition, self-worth, and wellbeing. The findings also suggest that adolescent wellbeing extends beyond stress reduction and includes self-acceptance, positive relationships, purpose, personal growth, and opportunities for rest. Overall, the study challenges simplistic East–West educational comparisons and highlights the importance of understanding adolescent wellbeing as a multidimensional and culturally embedded phenomenon.

Keywords adolescent mental wellbeing; Eastern-style schooling; Western-style schooling; parenting; autonomy; Shanghai

1.Introduction

Adolescent mental wellbeing has become an increasingly important concern in contemporary education, especially in large urban contexts where academic competition, family expectations, and globalized schooling models intersect. Shanghai provides a particularly meaningful setting for examining this issue because it contains both highly competitive Chinese public-school systems and international or Western-style educational environments. Adolescents growing up in this context are often exposed to different assumptions about academic success, classroom participation, teacher authority, parental involvement, autonomy, creativity, and future

planning. These different educational and family environments may shape how young people understand pressure, support, achievement, self-worth, and psychological wellbeing.

Discussions of Eastern-style and Western-style education often rely on simplified contrasts. Eastern-style schooling is commonly associated with discipline, examination performance, teacher authority, structured instruction, and strong family investment. Western-style schooling is often associated with student-centered learning, classroom discussion, creativity, self-expression, and personal autonomy. However, this binary distinction is insufficient. Chinese education has been recognized for strong academic outcomes, yet scholars have also argued that its high-stakes testing culture may constrain autonomy, creativity, deep learning, and socio-emotional development (Yu et al., 2018). At the same time, Western-style schooling should not be automatically understood as a low-pressure alternative. Although it may provide more opportunities for expression and exploration, it can also create pressure through GPA competition, extracurricular expectations, self-directed learning demands, social performance, and preparation for overseas higher education.

Parenting further complicates the relationship between schooling and adolescent wellbeing. In Chinese families, parenting is often closely connected to educational responsibility, moral guidance, emotional care, and long-term family expectations. Classical Western parenting models tend to distinguish between authoritative and authoritarian parenting, but these categories do not always capture the meanings of Chinese parental involvement (Chao, 1994, 2000). Chinese parents may express care through supervision, guidance, correction, educational investment, and practical support. Recent studies suggest that Chinese supportive parenting may combine high involvement, guidance, instruction, educational investment, and selective autonomy support rather than fitting neatly into Western categories of warmth and control (Zhang et al., 2024; Liu et al., 2025).

Therefore, adolescent mental wellbeing should not be examined only through the separate effects of school or family. Instead, it should be understood through the interaction between schooling style, parenting practice, and adolescents' own interpretation of pressure and support. A highly structured school environment may strengthen competence but reduce autonomy if students feel controlled by external expectations. A more flexible educational environment may support self-expression but also increase stress when students must manage unclear expectations or excessive self-directed work. Similarly, parental involvement may function as emotional support, practical care, pressure, or conditional approval depending on how it is communicated and experienced.

This qualitative study examines how Western-style and Eastern-style schooling and parenting shape adolescent mental wellbeing in Shanghai. Rather than asking which educational model is better, the study investigates how different school and family environments produce different combinations of academic pressure, autonomy, teacher support, parental involvement, emotional security, recognition, and self-development. By adopting a culturally sensitive and multidimensional understanding of wellbeing, this study seeks to move beyond simple East-West

comparisons and contribute to a more nuanced understanding of adolescent development in a globalized Chinese urban context.

2.Literature Review

2.1.Eastern-style schooling, achievement, and psychological pressure

Research on Chinese education has long emphasized the tension between academic achievement and developmental cost. Chinese students, particularly those in major cities such as Shanghai, have performed strongly in international academic assessments. This achievement has often been associated with structured instruction, disciplined study habits, high expectations, and strong family support. However, the same educational ecology has also been criticized for its reliance on high-stakes examinations, teacher-centered instruction, intensive homework, and narrow evaluation standards (Yu et al., 2018).

Yu et al. (2018) argue that self-determination theory provides a useful lens for understanding Chinese education because it explains how high academic performance may coexist with psychological strain. In such systems, students may develop competence through structure, repetition, and clear standards, but their autonomy may be constrained when learning is primarily directed by examinations, teachers, and parental expectations. Relatedness may also become ambivalent: students may feel supported by teachers and parents, while also fearing that poor performance will disappoint them.

Eastern-style schooling should not be interpreted only as a source of harm. Discipline, teacher guidance, and structured learning can help students develop persistence, academic competence, and a sense of direction. In Chinese cultural contexts, educational effort is often connected to family responsibility and long-term mobility. Students may understand academic success not merely as individual achievement but also as a response to parental sacrifice and social expectation. This means that academic pressure may be experienced together with motivation, purpose, and responsibility.

However, when examination performance becomes the dominant measure of student value, it may weaken adolescent wellbeing. A narrowly achievement-oriented environment can reduce opportunities for creativity, emotional expression, social exploration, and non-academic development. It may also encourage students to connect self-worth too closely with grades, rankings, or teachers' evaluations. This is especially important during adolescence, a developmental period in which identity, autonomy, peer relationships, and emotional regulation become increasingly significant. Therefore, the psychological effects of Eastern-style schooling depend not simply on discipline or examination culture itself, but on whether structure is combined with emotional support, teacher boundaries, recognition of diverse talents, and reasonable parental expectations.

2.2. Western-style schooling, autonomy, and hidden pressure

Western-style schooling is commonly associated with autonomy-supportive learning, active participation, creativity, communication, and holistic development. Compared with traditional teacher-centered instruction, Western educational philosophies often give students more opportunities to express opinions, participate in classroom discussion, explore interests, and take responsibility for their own learning (Hassan et al., 2010; Iyer, 2015). These features may benefit adolescent wellbeing by supporting autonomy, confidence, positive teacher-student relationships, and personal growth.

From the perspective of self-determination theory, Western-style schooling may be psychologically beneficial when it supports students' basic needs for autonomy, competence, and relatedness (Ryan & Deci, 2024). Autonomy-supportive classrooms can help students feel that learning is meaningful and self-endorsed. Discussion-based teaching may strengthen relatedness by allowing students to feel heard and respected. Project-based or inquiry-based learning may support personal growth by encouraging students to connect schoolwork with interests and real-world problems.

However, Western-style schooling should not be idealized as naturally stress-free. In international or Western-style schools, academic pressure may take different forms. Students may face GPA competition, college application pressure, extracurricular demands, leadership expectations, social comparison, and the need to appear confident, expressive, and well-rounded. In this sense, Western-style schooling may transform pressure rather than eliminate it. Instead of being judged only by test scores, students may be evaluated through a broader but still demanding set of criteria, including academic performance, creativity, social skills, extracurricular achievement, and future planning.

Autonomy may also become stressful when it is not accompanied by adequate structure and guidance. Self-directed learning can support growth when students receive clear expectations, timely feedback, and emotional support. However, when students are expected to manage heavy workloads, unclear instructions, or complex academic tasks largely on their own, autonomy may become a burden. Therefore, Western-style schooling may support wellbeing when autonomy is balanced with structure, but it may undermine wellbeing when independence becomes excessive responsibility.

This is particularly relevant in Shanghai, where international education is often linked to global competition and overseas university pathways. Families may choose Western-style schooling because they value creativity, English-language learning, global exposure, and holistic development. Yet these aspirations may also create new forms of pressure. Students may be expected to perform well academically while also building a distinctive personal profile. Thus, Western-style schooling should be analyzed as a complex educational environment that provides both resources and risks for adolescent wellbeing.

2.3.Parenting, culture, and adolescent wellbeing

Parenting plays a central role in adolescent mental wellbeing, but parenting must be understood within cultural context. Classical parenting theory often distinguishes among

authoritative, authoritarian, permissive, and neglectful styles. In many Western studies, authoritative parenting, which combines warmth and firm guidance, is associated with positive child outcomes, while authoritarian parenting, which combines high control with low warmth, is often associated with less favorable outcomes (Baumrind, 1971; Darling & Steinberg, 1993). However, this framework may not fully explain Chinese parenting.

Chao (1994, 2000) argues that Chinese parenting should be interpreted through culturally specific concepts such as training and *guan*. In Chinese contexts, parental monitoring, correction, instruction, and educational involvement may express care, responsibility, and moral commitment. What appears to be control from a Western perspective may be understood within Chinese families as guidance, sacrifice, and support. Therefore, applying Western parenting categories without cultural sensitivity may lead to misleading interpretations of Chinese parental behavior.

Recent studies have further challenged stereotypes of Chinese parenting as simply authoritarian or "tiger parenting." Zhang et al. (2024) found that Chinese supportive parenting may include high involvement, guidance, instruction, respect for autonomy, and freedom to explore. Liu et al. (2025) similarly show that contemporary urban Chinese parenting increasingly incorporates child-centered practices, including guiding, nurturing, supporting inclusive autonomy, and investing in education. However, these practices may still carry emotional complexity. Young people may appreciate parental support and investment while also feeling obligation, pressure, or guilt.

Parenting may influence adolescent wellbeing through several mechanisms. First, parental expectations can shape students' motivation and sense of responsibility. When expectations are communicated with trust and emotional support, they may help adolescents develop purpose and competence. When expectations are experienced as conditional approval, they may increase anxiety and weaken self-acceptance. Second, parent-child communication can reduce or intensify stress. Liu (2023) found that parental involvement and parent-child communication can buffer the negative relationship between mental unwellness and academic performance among Chinese middle school students. However, the benefits of communication depend on its quality. Communication focused only on grades may increase pressure, while communication that acknowledges emotional difficulties may support wellbeing. Third, parental involvement may affect autonomy. Low-pressure companionship may support adolescent independence, while excessive monitoring may reduce agency.

2.4.School-family relationship and adolescent wellbeing

Schooling and parenting are often studied separately, but adolescent wellbeing is shaped by their interaction. Schools communicate academic expectations, behavioral norms, and evaluation standards to families. Families then translate these expectations into daily routines, study supervision, extracurricular planning, tutoring decisions, and emotional responses. The school-family relationship therefore forms an important mesosystem in adolescent development (Bronfenbrenner, 1979).

In Eastern-style schooling, school-family communication may reinforce academic monitoring and examination preparation. This can create consistency and structure, but it may also make students feel that school pressure extends into home life. In Western-style or international schooling, school-family communication may appear more holistic, focusing on personal development, activities, and student interests. However, families may still interpret holistic development through competitive future planning, especially when international education is connected to overseas university admission. Therefore, school-family interaction may create pressure in both educational systems, though through different mechanisms.

Yamamoto et al. (2022) argue that parental involvement should be reconceptualized through a sociocultural model because home-based and school-based involvement reflect different cultural logics. This argument is useful for understanding Chinese families in globalized educational settings. Chinese parents may be deeply involved even when they do not participate visibly in school events. Their involvement may appear in arranging study schedules, selecting extracurricular activities, discussing future pathways, communicating with teachers, or providing emotional and financial support.

Thus, adolescent wellbeing depends not only on the amount of parental involvement or school support, but also on how these forms of involvement are experienced. Teacher strictness may be protective when it is combined with fairness, guidance, and respect for students' personal lives. Parental involvement may be supportive when it is expressed as companionship and trust, but stressful when it becomes surveillance or grade-centered judgment. This study therefore examines schooling and parenting as interconnected environments that shape adolescent wellbeing through pressure, support, autonomy, and recognition.

2.5.Research gap

Existing literature provides valuable insights into Chinese education, Western-style schooling, parenting culture, and adolescent wellbeing. However, three gaps remain. First, many studies examine school education or family parenting separately, while fewer studies analyze how schooling style, parenting practice, and school-family interaction jointly shape adolescent wellbeing. Second, much research on Chinese parenting focuses on immigrant or overseas Chinese families, while fewer studies examine adolescents living in contemporary Chinese cities where Eastern-style and Western-style education coexist. Third, existing discussions often rely on broad East-West contrasts, while fewer qualitative studies explore the more complex ways adolescents understand pressure, support, autonomy, recognition, and self-worth.

This study addresses these gaps by focusing on Shanghai adolescents in different educational settings. It does not assume that Western-style schooling is automatically healthier or that Eastern-style schooling is necessarily more harmful. Instead, it examines how each educational environment may generate both wellbeing resources and psychological risks. In doing so, the study contributes to a more balanced and culturally grounded understanding of adolescent mental wellbeing in contemporary China.

3.Theoretical Framework

This study adopts an integrated theoretical framework that combines the social ecological model, self-determination theory, culturally responsive parenting theory, and Ryff's model of psychological wellbeing. These perspectives are complementary. The social ecological model explains why adolescent wellbeing must be understood across interacting environments. Self-determination theory explains how schooling and parenting may support or frustrate basic psychological needs. Culturally responsive parenting theory explains why Chinese parenting should not be evaluated only through Western categories. Ryff's model provides a multidimensional understanding of wellbeing beyond stress reduction.

3.1.Social ecological model: adolescent wellbeing across school, family, and culture

The social ecological model, associated with Bronfenbrenner (1979), emphasizes that adolescent development occurs within nested and interacting systems. These include the family, school, peer group, community, cultural values, and broader social structures. From this perspective, adolescent mental wellbeing is not simply an individual psychological condition. It is produced through everyday interactions among students, parents, teachers, schools, and cultural expectations.

This model is especially suitable for the present study because the research focuses on the combined effects of schooling and parenting. The school environment forms one microsystem, including curriculum, classroom culture, teacher authority, peer competition, workload, evaluation methods, and opportunities for self-expression. The family forms another microsystem, including parental expectations, emotional support, communication style, educational investment, and autonomy support. The interaction between school and family forms the mesosystem, where school expectations are translated into parental practices and student experiences.

In Shanghai, these systems are shaped by broader cultural and social forces. Eastern-style schooling is influenced by Chinese values of diligence, discipline, educational mobility, respect for authority, and family responsibility. Western-style schooling is influenced by global ideals of independence, creativity, self-expression, and holistic development. However, these systems are not pure opposites. Public schools may include elements of holistic education, while international schools may still reproduce intense competition. Families may combine Chinese educational expectations with global aspirations. Therefore, the social ecological model allows this study to analyze adolescent wellbeing as a product of overlapping school, family, and cultural systems rather than a simple outcome of one educational model.

3.2.Self-determination theory: autonomy, competence, and relatedness

Self-determination theory provides the second foundation of the framework. According to Deci and Ryan (2000) and Ryan and Deci (2024), human motivation and wellbeing depend on the satisfaction of three basic psychological needs: autonomy, competence, and relatedness. Autonomy refers to the feeling that one's actions are self-endorsed. Competence refers to the

feeling that one can master tasks and achieve meaningful goals. Relatedness refers to the feeling of being cared for, understood, and connected to others.

Yu et al. (2018) apply self-determination theory to Chinese education and argue that it helps explain the paradox of strong academic achievement and psychological cost. Chinese education may support competence through structure, effort, and high standards. However, it may undermine autonomy when students feel controlled by examinations, teacher authority, and parental expectations. Relatedness may also become ambivalent when support is tied closely to performance.

Self-determination theory helps this study compare the psychological mechanisms of Eastern-style and Western-style schooling. Eastern-style schooling may strengthen competence by providing clear goals, disciplined instruction, and structured feedback. However, it may weaken autonomy when students have limited choice or few opportunities to express views. Western-style schooling may support autonomy through discussion, exploration, and flexible learning, but it may challenge competence when students face unclear expectations or excessive self-directed tasks. Both systems may either support or weaken relatedness depending on the quality of teacher-student and parent-child relationships.

A key point in this study is that autonomy is not automatically beneficial. Autonomy supports wellbeing when it is accompanied by structure, guidance, and emotional support. Without these conditions, autonomy may become a source of stress. For adolescents, freedom to explore must be balanced with clear instruction, manageable workload, and reliable support. Therefore, this study treats autonomy as a conditional resource rather than a simple marker of educational superiority.

3.3.Culturally responsive parenting theory: beyond authoritarian and authoritative categories

The third component of the framework is culturally responsive parenting theory. Traditional Western parenting models often classify parenting according to warmth and control. Authoritative parenting is typically understood as high warmth and firm guidance, while authoritarian parenting is understood as high control and low warmth (Baumrind, 1971; Darling & Steinberg, 1993). However, this model may misread Chinese parenting if it assumes that control and care are always opposite.

Chao (1994, 2000) argues that Chinese parenting should be understood through culturally specific meanings of training and *guan*. In Chinese families, parental guidance, correction, monitoring, and educational involvement may express care, responsibility, and moral commitment. This does not mean that all strict parenting is beneficial, but it does mean that parenting practices must be interpreted through cultural meanings and adolescent perceptions.

Recent research supports this more nuanced view. Zhang et al. (2024) show that Chinese supportive parenting may include high involvement, instruction, guidance, autonomy, respect, and freedom to explore. Liu et al. (2025) find that contemporary urban Chinese parenting combines individualistic goals of child development with collectivistic practices of family

investment and responsibility. These studies suggest that Chinese parenting is not simply authoritarian; it is often a hybrid form shaped by tradition, social change, educational competition, and child-centered ideals.

For this study, culturally responsive parenting theory helps analyze how parental involvement may function as both support and pressure. Parental supervision may help adolescents develop competence, but it may also reduce autonomy if experienced as distrust. Educational investment may communicate care, but it may also create obligation. Low parental intervention may support independence, but it may also become problematic if students lack emotional support. Therefore, the study examines not only what parents do, but how adolescents interpret parental behavior in relation to trust, pressure, companionship, and self-worth.

3.4. Ryff's model of psychological wellbeing: beyond the absence of stress

The final component of the theoretical framework is Ryff's model of psychological wellbeing. Ryff and Keyes (1995) propose six dimensions of wellbeing: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Ryff (2013) further argues that wellbeing should not be reduced to happiness, life satisfaction, or the absence of distress. Rather, wellbeing includes meaning-making, self-realization, resilience, and the ability to manage life challenges.

This model is appropriate because adolescent wellbeing in educational contexts is multidimensional. A student may experience academic stress but still feel purposeful, capable, and supported. Another student may have more freedom but feel uncertain, socially exhausted, or insufficiently guided. Ryff's framework allows this study to examine how schooling and parenting shape not only stress levels but also identity, self-worth, growth, relationships, and meaning.

Autonomy in Ryff's model concerns students' ability to make choices and act according to personal values. Environmental mastery refers to students' ability to manage schoolwork, family expectations, extracurricular activities, and emotional demands. Personal growth refers to whether school and family help adolescents develop abilities beyond grades. Positive relations include trust, emotional communication, teacher support, parental companionship, and peer connection. Purpose in life concerns whether students understand education as meaningful or merely compulsory. Self-acceptance refers to whether students can value themselves beyond rankings, grades, and external approval.

This model is especially useful for avoiding a narrow definition of mental wellbeing. In education, wellbeing is often discussed as the reduction of anxiety or stress. However, adolescents also need recognition, purpose, self-understanding, and opportunities to develop diverse strengths. A school environment may reduce pressure but fail to provide purpose. A family environment may provide resources but weaken self-acceptance if approval depends only on achievement. Therefore, Ryff's model helps this study analyze adolescent wellbeing as a broader developmental condition.

3.5.Integrated analytical framework

Taken together, this study conceptualizes adolescent mental wellbeing as the result of interactions among schooling style, parenting practice, and adolescents' interpretation of pressure and support. Eastern-style and Western-style schooling are not treated as fixed opposites. Instead, they are understood as educational environments that organize structure, autonomy, competition, recognition, and teacher-student relations in different ways. Parenting practices are also understood as culturally embedded forms of guidance, companionship, monitoring, trust, and expectation.

The integrated framework guides the analysis through four key concepts. First, perceived pressure includes examination pressure, GPA competition, workload, extracurricular demands, social comparison, and future-planning anxiety. Second, perceived support includes teacher guidance, parental companionship, emotional communication, peer connection, and recognition of non-academic strengths. Third, autonomy and agency refer to students' opportunities to express opinions, explore interests, make choices, and manage learning with sufficient support. Fourth, psychological wellbeing includes self-acceptance, environmental mastery, personal growth, positive relations, and purpose.

This framework is suitable for a qualitative study because it allows adolescent experiences to be understood in context. It avoids assuming that Western-style education is naturally liberating or that Eastern-style education is necessarily oppressive. Instead, it examines how each system may produce both resources and risks. In Shanghai, where global educational ideals and Chinese family expectations intersect, such an approach offers a balanced and culturally grounded way to understand adolescent mental wellbeing.

4. Methods

4.1 Research Design

This study adopted a qualitative research design to explore how adolescents in Shanghai understand the effects of Eastern-style and Western-style schooling and parenting on their mental wellbeing. Because the study aimed to examine students' subjective experiences of academic pressure, parental involvement, autonomy, emotional support, teacher-student relationships, and self-development, a qualitative approach was considered appropriate.

The study was based on 23 semi-structured in-depth interview materials collected from adolescent students studying in either Eastern-style or Western-style educational settings. These interviews allowed participants to describe their school experiences, family relationships, perceptions of academic pressure, and emotional wellbeing in their own words. Rather than measuring wellbeing through predetermined psychological scales, the study focused on how students themselves interpreted pressure, support, autonomy, achievement, and self-worth.

The research compared adolescents who had experience with Eastern-style education, represented mainly by the Chinese domestic school system, and those who had experience with Western-style education, represented mainly by international, bilingual, or Western-oriented

schooling. However, the study did not treat these two educational systems as fixed opposites. Instead, it examined how students in both systems experienced different combinations of structure, competition, freedom, teacher support, parental expectation, and emotional security.

4.2 Participants and Sampling

The participants consisted of 23 adolescent students from middle school and high school levels. They were studying in different educational settings, including traditional Chinese schools, international schools, Western-style schools, and Western-oriented educational programs. The sample included students from Grade 6 to Grade 12, allowing the study to capture experiences across early adolescence, middle adolescence, and late adolescence.

Participants were recruited through convenience sampling and snowball sampling. The researcher first contacted students who were accessible through existing social networks, including school-related and extracurricular contexts. Some participants then helped introduce or distribute the study to other students with similar educational backgrounds. This recruitment strategy was suitable because the study focused on adolescents' personal experiences and required participants who were willing to reflect on school, family, and wellbeing in some detail.

The final dataset contained 23 interview materials. These materials included students from both Eastern-style and Western-style schooling contexts. Eastern-style participants were mainly students in the Chinese domestic education system, where schooling is usually characterized by structured instruction, examination pressure, teacher authority, homework intensity, and close school-family communication. Western-style participants were mainly students in international or Western-oriented schools, where schooling tends to include more classroom discussion, extracurricular activities, GPA-based evaluation, self-directed learning, and preparation for overseas education.

Although the participants were grouped broadly according to schooling style, the analysis did not assume that all Eastern-style students shared the same experience or that all Western-style students experienced schooling in the same way. Instead, the study paid attention to internal differences within each group, including differences in grade level, teacher support, parental involvement, school culture, and students' own coping styles.

4.3 Data Collection

Data were collected through 23 semi-structured in-depth interviews or interview-style open-ended responses. The interviews focused on students' lived experiences of schooling, parenting, academic pressure, autonomy, teacher support, parent-child communication, and mental wellbeing.

Before participating, students were informed about the purpose of the study and the voluntary nature of participation. They were told that the study aimed to understand how different school and family environments shaped adolescent wellbeing, rather than to judge whether one educational system was better than another. Participants were encouraged to answer based on their own experiences and were allowed to skip questions they did not want to answer.

The interviews followed a semi-structured format. This means that the researcher used a set of guiding questions, but participants were also allowed to elaborate freely. The semi-structured format was useful because it provided consistency across participants while still allowing individual experiences to emerge. When participants mentioned issues such as examination pressure, GPA competition, parental control, teacher support, social exhaustion, sleep problems, or lack of autonomy, these responses were treated as important qualitative evidence for further analysis.

The interview questions covered several key areas. Participants were asked to describe their school life, their feelings about academic pressure, their relationship with teachers, their opportunities to express opinions or explore knowledge, their parents' involvement in education, their feelings about parent-teacher communication, and the changes they hoped would make school life healthier or happier. These questions were directly connected to the study's theoretical framework, especially perceived pressure, perceived support, autonomy and agency, and psychological wellbeing.

4.4 Research Instrument

The interview guide was designed according to the study's integrated theoretical framework. It drew on the social ecological model, self-determination theory, culturally responsive parenting theory, and Ryff's model of psychological wellbeing. The questions focused on four main areas: perceived academic pressure, perceived support, autonomy and agency, and psychological wellbeing.

First, questions about perceived pressure explored students' experiences of homework, examinations, GPA competition, extracurricular expectations, social comparison, future planning, and school schedules. These questions helped identify how pressure appeared differently in Eastern-style and Western-style schooling.

Second, questions about perceived support examined teacher guidance, parental companionship, parent-child communication, peer relationships, and emotional recognition. These questions were used to understand whether students experienced school and family involvement as supportive, controlling, or emotionally distant.

Third, questions about autonomy and agency asked whether students felt they had opportunities to express opinions, explore personal interests, make choices, and manage learning independently. This was important because the study did not assume that autonomy was always positive. Instead, it examined whether autonomy was accompanied by sufficient structure, guidance, and emotional support.

Fourth, questions about psychological wellbeing explored students' feelings of self-acceptance, personal growth, environmental mastery, positive relationships, purpose in education, tiredness, anxiety, pressure, and recognition. These questions reflected the study's broader understanding of wellbeing as more than the absence of stress.

The open-ended format was used because the study aimed to understand students' own interpretations rather than force their experiences into fixed categories. Participants were

encouraged to describe their experiences in their own words and to provide examples when possible.

4.5 Ethical Considerations

Ethical considerations were addressed throughout the research process. Before participating, students were informed of the purpose of the study, the voluntary nature of participation, and their right to stop participating or refuse to answer any question. The researcher avoided pressuring participants to disclose information they were uncomfortable sharing.

Because the study involved adolescents and topics related to mental wellbeing, the questions were framed carefully around school experience, family support, academic pressure, emotional wellbeing, and everyday life rather than clinical diagnosis or personal trauma. When participants expressed negative emotions, stress, anxiety, or dissatisfaction, these responses were treated respectfully and were not interpreted as clinical evidence.

Participants' privacy was protected by using respondent numbers rather than real names. Identifying information was removed from the analysis. The study also avoided including details that could reveal a participant's specific school, family identity, or personal background. The purpose of presenting participant quotations was to illustrate broader themes, not to expose individual experiences.

One participant's responses were very brief and ambiguous, making them difficult to interpret meaningfully. This material was still counted as part of the collected dataset, but it was not used as core evidence in the thematic analysis. This decision helped preserve the overall sample record while maintaining the reliability of the qualitative interpretation.

4.6 Data Analysis

The data were analyzed using thematic analysis. All 23 interview materials were reviewed carefully to identify repeated patterns, meaningful expressions, and differences between students from Eastern-style and Western-style educational backgrounds. The analysis was guided by the study's integrated theoretical framework, including the social ecological model, self-determination theory, culturally responsive parenting theory, and Ryff's model of psychological wellbeing.

The first stage of analysis involved reading the interview materials multiple times to become familiar with the data. During this stage, attention was paid to how students described school life, academic pressure, teacher support, parental involvement, autonomy, emotional security, fatigue, anxiety, self-worth, and achievement.

The second stage involved generating initial codes. These codes included examination pressure, GPA competition, homework burden, extracurricular pressure, social exhaustion, passive learning, active exploration, teacher guidance, teacher criticism, parental control, parental companionship, parent-teacher communication, lack of trust, emotional recognition, rest, sleep, self-acceptance, and non-academic development.

The third stage involved grouping these codes into broader themes. The main themes included different forms of academic pressure, autonomy as both support and burden, parenting as both care and control, parent-teacher communication as both support network and pressure mechanism, teacher support as guidance with emotional recognition, and psychological wellbeing as recognition, rest, purpose, and self-acceptance.

The final stage involved comparing patterns across Eastern-style and Western-style schooling contexts. The analysis did not aim to prove that one system was better than the other. Instead, it examined how each educational environment produced different combinations of structure, competition, support, freedom, responsibility, and emotional experience. This approach allowed the study to move beyond a simple East-West binary and develop a more culturally grounded understanding of adolescent mental wellbeing in Shanghai.

5. Findings

This study conducted a thematic analysis of interview and open-ended questionnaire data to examine Shanghai adolescents' learning experiences, family upbringing experiences, and perceptions of psychological wellbeing within both Eastern-style schooling and Western-style schooling contexts. The findings indicate that students' perceptions of school and family cannot be simply summarized as "Eastern education is more stressful" or "Western education is more liberating." Rather, both educational environments can simultaneously generate support and pressure, although the forms of support and pressure differ.

Students in Eastern-style schooling more frequently mentioned examination frequency, homework burden, passive learning, tensions arising from school–family communication, and restrictions imposed by daily rules. In contrast, students in Western-style schooling more often referred to GPA competition, extracurricular workload, pressure to perform socially, self-management demands, and the hidden competition associated with the ideal of "holistic development."

Among all interview and questionnaire materials, the responses provided by Respondent 22 were extremely brief and ambiguous, consisting largely of answers such as ".", "at," "no," and "this," which could not be meaningfully interpreted. Therefore, these responses were considered low-quality data and were not used as primary evidence in the core thematic analysis.

5.1 Different Educational Systems Produce Different Forms of Pressure, but Pressure Does Not Disappear in Either System

The first major finding is that both Eastern-style schooling and Western-style schooling generate pressure; however, the sources, manifestations, and interpretations of that pressure differ significantly.

In Eastern-style schooling, pressure primarily originates from examinations, homework, daily discipline, and passive learning. Some students described school life as highly repetitive, stressful, and even painful. For example, Respondent 5, a ninth-grade female student in a traditional Chinese school, described her school life as "rule-bound, nerve-racking, and worse

than death." She explained that the constant cycle of classes and examinations generated extremely negative emotions whenever she performed poorly. This suggests that for some students in Eastern-style schooling, examinations function not only as evaluation mechanisms but also as major determinants of self-worth and emotional stability.

Similarly, Respondent 11 noted that while school life included "rich activities" and "harmonious teacher–student relationships," it was also "academically demanding." She explicitly expressed a desire for "less homework, more activities, and fewer examinations." This indicates that although Eastern-style schooling is not devoid of extracurricular activities or supportive teacher–student relationships, academic workload and examination intensity continue to significantly affect students' daily wellbeing.

Respondent 22 also hoped that schools would "eliminate evening self-study sessions, reduce examination frequency, and organize more outings." This suggests that students' stress stems not only from examinations themselves but also from structural factors such as school schedules, mandatory evening study programs, and insufficient opportunities for relaxation.

By contrast, pressure in Western-style schooling is less likely to appear as a single examination-based burden and more likely to emerge through GPA competition, extracurricular commitments, social expectations, university preparation, and self-presentation demands. Respondent 2, a ninth-grade female student attending an American boarding school, described her school life as "exhausting, intensive, but socially positive." Her comments suggest that although international or Western-style schools provide abundant activities and social opportunities, this richness itself may create fatigue and time pressure. She also emphasized the importance of GPA, indicating that these schools are not necessarily "easy" environments.

Respondent 3 articulated the pressure of Western-style schooling even more directly. She summarized school life simply as "stressful" and argued that although schools claim not to focus exclusively on grades, pressure is still created through grading systems, homework, and examination schedules. She also noted that many major assessments were scheduled on the same day, forcing students to spend entire weekends completing assignments. This demonstrates that pressure in Western-style schooling may not take the form of traditional test-oriented education but instead manifests through multitasking demands, time management challenges, and GPA-based evaluation systems.

Respondent 6 similarly described her international school as "diverse, competitive, and somewhat involuted." She expressed a desire for schools to resist excessive competition and avoid placing academic achievement above everything else. She further observed that some students possessed exceptional athletic or artistic talents, yet the school culture's emphasis on grades suppressed young people's dreams. This suggests that although Western-style schooling promotes holistic development in principle, students may still experience it as another form of competitive culture.

Therefore, the findings indicate that the two educational systems should not be viewed as a simple contrast between "high pressure" and "low pressure." Pressure in Eastern-style schooling is concentrated around examinations, homework, discipline, and university admission

pathways, whereas pressure in Western-style schooling is more diffuse, appearing through GPA requirements, extracurricular achievements, social performance, self-management, and future planning. Western-style schooling does not eliminate pressure; rather, it transforms its form.

5.2 Autonomy Is Both a Psychological Resource and a Potential Burden

The second major finding is that autonomy does not automatically enhance psychological wellbeing. Whether students benefit from autonomy depends on whether it is accompanied by clear guidance, reasonable workloads, and emotional support.

In Western-style schooling, many students reported having opportunities to express their opinions and actively explore knowledge. For example, Respondent 1 believed that classroom learning and extracurricular exploration were well balanced. He felt that he could learn from teachers while also independently pursuing knowledge beyond the classroom, describing this model as "good." This suggests that when independent exploration is balanced with teacher guidance, students are more likely to experience autonomy as an opportunity for growth.

Respondent 2 likewise stated that she had sufficient opportunities to express her views and felt that active exploration provided her with "the freedom to develop her own ideas." Respondent 8 noted that teachers frequently encouraged students to think through questioning, making classroom participation and self-expression positive learning experiences.

However, autonomy in Western-style schooling can also become burdensome. Respondent 3 explained that although students enjoy considerable freedom to explore, there are times when they cannot fully understand complex concepts on their own. In such situations, being expected to "self-study" becomes frustrating and difficult. This suggests that when schools interpret independent learning as students solving problems entirely by themselves, without adequate instructional support, autonomy may become a source of anxiety and helplessness.

Respondent 7's comments further reveal the exhausting side of autonomy. She stated that she was "too tired" to explore knowledge beyond what was necessary and therefore often chose passive learning instead. She summarized her condition as "sleep, stress, and annoyance" and argued that some assignments were unnecessary. Her experience demonstrates that even in Western-style schooling, students may lose the motivation to engage in active exploration when workloads become overwhelming.

5.3 The Impact of Parental Involvement Depends on Whether Students Experience Trust, Companionship, and Respect for Boundaries

The third major finding is that parental involvement is neither inherently supportive nor inherently stressful. How students interpret parental involvement depends on whether it is experienced as trust, companionship, and guidance, or as control, surveillance, and conditional support based on academic performance.

Some students viewed relatively low levels of parental intervention as a form of positive support. Respondent 1 explained that as long as his academic performance remained acceptable, his parents did not interfere excessively, and he appreciated this arrangement. Respondent 4

similarly noted that his parents were not highly involved in his schooling, and he found this relatively free environment comfortable.

Respondent 6 described a family support pattern characterized by low intervention but strong emotional companionship. She stated that her parents were not heavily involved in her academic life, yet they provided companionship and did not evaluate her solely based on grades. As a result, she felt relatively free and secure. This suggests that low parental intervention does not necessarily imply a lack of support; when parents provide emotional companionship and recognition beyond academic achievement, adolescents may experience greater autonomy and psychological safety.

In contrast, when parental involvement is perceived as controlling, distrustful, or conditional, students are more likely to experience stress. Respondent 5 reported that although her parents were not highly involved, their strictness and lack of trust made her feel pressured and constrained. In this case, the issue was not the amount of parental involvement but rather the way that involvement communicated distrust.

Respondent 21 provided a particularly clear example of dissatisfaction resulting from high parental control. She described her parents as highly involved but granting very little freedom, a situation she explicitly disliked. She expressed a desire for greater autonomy in basic aspects of daily life, including "the freedom to drink water, the freedom to use the restroom, and the right to express my feelings." These comments suggest that for some students, psychological wellbeing begins with basic personal freedoms and opportunities for self-expression, rather than solely with academic achievement or educational success.

Among students in Western-style schooling, parental pressure had not disappeared but often manifested through academic expectations and conditional support. Respondent 7 stated that if she failed to rank among the top ten students in her class, her mother would likely enroll her in numerous tutoring programs. Although she regarded this response as somewhat acceptable, she also expressed discomfort with being constantly monitored. This reflects a form of conditional support in which parents may not intervene heavily in daily life but increase control when academic performance falls below expectations.

Respondent 23, another student in Western-style schooling, reported that although his parents were not particularly strict, they still held expectations regarding academic achievement. He felt happy when his hard work resulted in good grades. This suggests that even within relatively relaxed family environments, academic performance remains an important factor influencing parent–child relationships and self-evaluation.

Overall, the findings suggest that the impact of parental involvement on adolescent wellbeing depends primarily on three factors: whether parents treat grades as the sole standard of evaluation, whether they provide trust and emotional companionship, and whether they respect adolescents' personal boundaries. Supportive parental involvement is not defined by either high or low levels of monitoring but by whether students feel understood, trusted, and accepted.

5.4 School–Family Communication Can Function as Both a Support Network and a Source of Pressure

The fourth major finding is that school–family communication plays a dual role in students' experiences. On the one hand, communication between teachers and parents can create a valuable support network. On the other hand, it can also make students feel monitored, anxious, or subject to increased pressure.

In several positive cases, students perceived school–family communication as helpful. Respondent 10 explained that when teachers communicated with her parents about her situation, she listened carefully and accepted their feedback because she believed both teachers and parents were trying to help her. This suggests that when students perceive school–family communication as supportive rather than punitive, it can enhance feelings of security and direction.

Respondent 8 similarly believed that although communication between her parents and teachers was not frequent, it helped her parents better understand the difficulties she faced. This indicates that moderate communication may improve parent–child relationships by increasing parental awareness of students' challenges and pressures.

However, many students in Eastern-style schooling associated school–family communication with feelings of tension. When asked about communication between parents and teachers, Respondent 11 simply responded, "nervous." Respondent 13 similarly reported feeling anxious whenever such communication occurred. Respondent 22 described school–family communication as "acceptable" but nevertheless associated it with tension. These responses suggest that within traditional school settings, students often interpret communication between teachers and parents as a mechanism for evaluation, supervision, or problem reporting rather than as a source of support.

Respondent 23 expressed an even stronger negative reaction, describing school–family communication as "uncomfortable." She also characterized her learning experience as "passive and unpleasant" and hoped that schools would demonstrate greater respect for students and respond more effectively to their needs. Her comments suggest that when students' voices are excluded from school–family communication, they may feel that they are merely objects of evaluation and management rather than individuals whose perspectives are valued.

Therefore, the findings indicate that school–family communication itself is not problematic. Rather, its impact depends on whether students' perspectives are included in the process. When communication focuses solely on grades, discipline, and problem reporting, it may extend school-related pressure into the home environment. When communication helps parents understand students' challenges and facilitates collaborative support, it can serve as a valuable protective resource.

5.5 Students Expect Teacher Support to Involve Understanding, Guidance, and Appropriate Boundaries Rather Than Mere Leniency

The fifth major finding is that students' understanding of a "good teacher" or a "supportive teacher" does not simply involve leniency. Many students hope that teachers can balance strictness, guidance, understanding, and respect for students' personal lives.

Among students in Eastern-style schooling, some accepted or even preferred strict teachers, provided that such strictness did not excessively intrude into their personal lives. Respondent 4 argued that an ideal teacher should "maintain discipline but not interfere in students' personal lives." He believed that classroom instruction could be intensive and demanding, but students should not be made to feel exhausted in their private lives as a result. This suggests that students do not necessarily reject discipline or high expectations; rather, they object to boundaryless management and chronic exhaustion.

Respondent 14 similarly described supportive teachers as those who are "serious and strict." This indicates that some students regard strictness as a legitimate aspect of a teacher's professional responsibility, provided it serves learning rather than humiliating or oppressing students.

At the same time, many students emphasized the importance of understanding and emotional support. Respondent 22 expressed a desire for teachers who would "understand me and not criticize me simply because I performed poorly on an exam." This suggests that students do not necessarily seek lower standards; rather, they wish to be respected and understood even when their academic performance falls short of expectations.

Respondent 21 believed that good teachers should genuinely care about students, while Respondent 23 emphasized the importance of teachers who encourage learners. Together, these responses point to an important finding: students expect teachers not only to transmit knowledge but also to provide emotional recognition and non-humiliating support.

Students in Western-style schooling placed greater emphasis on guidance, inspiration, and an engaging classroom atmosphere. Respondent 8 highlighted how teachers encouraged critical thinking through questioning. Respondent 15 described supportive teachers as "humorous and interesting," while Respondent 10 emphasized teachers who present ideas and encourage students to explore them independently. These responses suggest that teacher support is not merely about helping students achieve higher grades; it also involves stimulating curiosity, creating opportunities for self-expression, and reducing feelings of classroom oppression.

Overall, the findings indicate that students expect teacher support to possess three key characteristics: clear academic guidance, emotional understanding and encouragement, and appropriate boundaries that prevent poor performance from becoming a source of humiliation or excessive intervention.

5.6 Psychological Wellbeing Involves More Than Reduced Stress: It Also Includes Recognition, Rest, Purpose, and Self-Acceptance

The sixth major finding is that students do not define psychological wellbeing simply as the absence of pressure. They also care about whether they are recognized, whether they have opportunities to rest, whether they can maintain a healthy life rhythm, whether they can pursue

interests beyond academics, and whether they can separate their self-worth from academic performance.

Several students explicitly linked feelings of achievement to academic success. Respondent 5 stated that performing well in physics generated a sense of accomplishment, whereas poor performance created stress. Respondent 10 similarly reported feeling accomplished when her efforts resulted in satisfactory grades. Respondent 23 noted that achieving good results after working hard made him feel happy. These responses suggest that academic achievement remains an important source of psychological wellbeing because it provides feelings of competence, validation of effort, and external recognition.

However, when grades become the sole source of self-worth, psychological vulnerability may increase. Respondent 6 criticized the belief that "grades are everything," arguing that such a mindset suppresses students' athletic and artistic talents. She expressed a desire for schools to reduce excessive competition and recognize diverse abilities. This indicates that students seek not only academic recognition but also acknowledgment of a broader range of talents and interests.

Rest and physical wellbeing also emerged as important dimensions of psychological wellbeing. Respondent 17 hoped that school would begin at 9:00 a.m. because early mornings made it difficult to wake up. She also wished for a longer lunch break because the existing break lasted less than an hour and was often consumed by eating and conversations with teachers. Respondent 13 suggested moving "homework-free days" to Fridays. These responses indicate that psychological wellbeing is closely connected to daily schedules, sleep, opportunities for rest, and recovery time.

Some students in Western-style schooling associated wellbeing challenges with social performance and the need to remain constantly engaged. Respondent 8 described school life as "overwhelming," characterized by constant social participation and abundant activities. Her comments suggest that the richness of opportunities available in international schools may also contribute to social exhaustion.

At the same time, several students reported high levels of satisfaction with their educational environments. Respondent 15 described her school as "relaxed," while Respondent 16 praised the school atmosphere, cleanliness, and teacher quality and stated that there was little she would change. These examples suggest that when school environments, teacher support, and students' personal needs align effectively, adolescents may experience high levels of comfort, stability, and wellbeing.

Overall, the findings demonstrate that adolescent psychological wellbeing is multidimensional. It involves not only the reduction of stress but also feelings of respect, understanding, recognition, opportunities for rest and self-expression, the development of non-academic interests, and the ability to establish self-worth beyond grades.

5.7 Significant Variation Exists Within Both Eastern-Style and Western-Style Schooling

The final major finding is that substantial variation exists within both Eastern-style and Western-style schooling. Consequently, neither educational system should be understood as a uniform model.

Within Eastern-style schooling, students reported experiences ranging from severe distress and suppression to adaptation, comfort, and positive growth. Respondent 5's account reflected intense examination pressure and psychological burden, whereas Respondent 10 described herself as well adapted to a structured and fast-paced learning environment and capable of experiencing exploration and achievement within a traditional school setting. Respondent 4 characterized his school life as "positive, relatively relaxed, and full of hope." As a student specializing in competitive swimming, he reported receiving encouragement rather than discrimination and believed that his school supported holistic development.

Similarly, Western-style schooling encompassed both freedom and relaxation as well as strong competition and fatigue. Respondent 1 appreciated that teachers did not pressure students to pursue high grades in the same way as traditional educational systems and felt able to balance classroom learning with independent exploration. In contrast, Respondents 3 and 7 emphasized stress, exhaustion, heavy workloads, examination scheduling, and the burden of self-directed learning.

These findings suggest that adolescent psychological wellbeing is not determined solely by whether an educational environment is labeled "Eastern" or "Western." Rather, wellbeing is shaped by the interaction of specific school policies, teacher support, parental involvement, peer relationships, workload, time management structures, and students' individual coping strategies.

For some students, structured environments provide security and direction; for others, the same structures produce feelings of control and suppression. Likewise, autonomous learning may foster freedom and personal growth for some adolescents but create confusion and additional burdens for others.

6. Discussion and Conclusion

This study examined how Eastern-style and Western-style schooling and parenting shape adolescents' mental wellbeing in Shanghai. Drawing on interview and open-ended questionnaire data from 23 adolescent students, the study found that adolescent wellbeing is not determined by schooling style alone. Rather, it is constructed through the interaction of school environments, family contexts, teacher support, parental involvement, and students' own interpretations of pressure, autonomy, recognition, and self-worth.

The findings challenge the simplified assumption that Eastern-style schooling is necessarily more stressful while Western-style schooling is naturally more supportive of wellbeing. Eastern-style schooling tends to generate more direct and visible forms of pressure through examinations, homework, discipline, strict schedules, and school–family communication. However, it may also provide students with structure, direction, academic competence, and a sense of stability when strictness is combined with teacher guidance and emotional support. In contrast, Western-style schooling is more likely to offer opportunities for

self-expression, classroom participation, extracurricular exploration, and personal autonomy. Nevertheless, it may also produce hidden pressures through GPA competition, extracurricular portfolios, social performance expectations, self-directed learning demands, and future planning for overseas education. Therefore, Western-style schooling does not eliminate pressure; it transforms pressure into more individualized and diffuse forms.

A key contribution of this study is its emphasis on students' subjective interpretation of pressure and support. The same educational practice may be experienced differently by different students. For some adolescents, structured learning and teacher strictness provide security and motivation; for others, they create control and emotional burden. Similarly, autonomy can support personal growth when accompanied by clear guidance, manageable workload, and emotional support, but it can become stressful when students are expected to manage complex tasks without sufficient help. This finding supports the view that autonomy should not be understood simply as freedom from control. In adolescent education, autonomy is most beneficial when it is balanced with structure, feedback, and reliable support.

The study also shows that parental involvement plays a complex role in adolescent wellbeing. Supportive parenting is not defined simply by how much parents participate in students' education. Rather, it depends on whether parental involvement communicates trust, companionship, emotional recognition, and respect for boundaries. When parents focus only on grades or use academic achievement as the condition for approval, adolescents may experience anxiety, guilt, or reduced self-acceptance. However, when parents provide companionship, practical support, and recognition beyond academic performance, their involvement may become an important protective factor. This finding suggests that Chinese parenting should not be interpreted only through Western categories such as authoritarian or authoritative parenting. Instead, it should be understood in relation to cultural meanings, family expectations, and adolescents' own perceptions.

Teacher support and school–family communication also emerged as important factors influencing wellbeing. Students did not simply expect teachers to be lenient. Many students accepted strictness when it was accompanied by fairness, guidance, encouragement, and respect for personal boundaries. Supportive teachers were described as those who could provide clear academic instruction while also recognizing students' emotional needs. Similarly, school–family communication was not inherently harmful or helpful. It became supportive when it helped parents understand students' difficulties and created a shared support network. However, it became stressful when students experienced it as surveillance, evaluation, or problem reporting. This suggests that students' voices should be included more actively in communication between schools and families.

Theoretically, this study contributes to research on adolescent wellbeing by integrating schooling, parenting, and students' subjective meaning-making within one analytical framework. The findings support the social ecological view that adolescent wellbeing is produced through interactions among family, school, and broader cultural expectations. They also extend self-determination theory by showing that autonomy, competence, and relatedness are shaped

differently across Eastern-style and Western-style educational environments. In addition, the findings support culturally responsive understandings of parenting by demonstrating that parental guidance, control, investment, and care cannot be evaluated without considering cultural context and student interpretation. Finally, by drawing on Ryff's model of psychological wellbeing, this study shows that adolescent wellbeing includes not only the reduction of stress but also self-acceptance, positive relationships, purpose, rest, recognition, and opportunities for personal growth.

Practically, the findings suggest that schools and families should move beyond the question of which educational model is better. The more important question is how each educational environment can reduce harmful pressure while preserving its strengths. Eastern-style schools may need to create more space for student voice, emotional expression, rest, and non-academic development while maintaining useful structure and academic guidance. Western-style schools may need to provide clearer expectations, stronger academic guidance, and better workload management so that autonomy does not become excessive responsibility. Parents should be encouraged to communicate expectations in ways that emphasize trust and support rather than conditional approval. Teachers should be supported in developing practices that combine academic standards with emotional understanding and respect for students' personal boundaries.

Despite its contributions, this study has several limitations. First, the study was based on a relatively small qualitative sample of 23 adolescents in Shanghai. Although the data provided rich insights into students' lived experiences, the findings cannot be generalized to all adolescents in China. Shanghai is a highly developed and internationally connected urban context, and students in other regions may experience different school cultures, family expectations, and social conditions. Future research could include participants from other cities, rural areas, and different socioeconomic backgrounds to examine whether similar patterns appear in other educational contexts.

Second, the study relied mainly on adolescents' self-reported experiences. While students' perspectives are essential for understanding wellbeing, parents, teachers, school counselors, and administrators may offer different interpretations of the same issues. Future studies could adopt a multi-perspective design by including interviews with parents and teachers in order to better understand how school–family interaction shapes adolescent wellbeing from different viewpoints.

Third, this study compared Eastern-style and Western-style schooling as broad categories. However, contemporary educational settings are increasingly hybrid. Some Chinese public schools have introduced student-centered teaching methods, while some international or Western-oriented schools remain highly competitive and achievement-driven. Future research should move beyond the East–West binary and examine more specific educational features, such as workload, teacher feedback, classroom participation, assessment style, extracurricular expectations, and school–family communication patterns.

Fourth, this study adopted a cross-sectional qualitative design and captured students' experiences at one point in time. Adolescents' perceptions of pressure, support, autonomy, and wellbeing may change across different developmental stages and school transitions. Future research could use longitudinal designs to explore how students' wellbeing changes over time, especially during key transitions such as moving from middle school to high school, entering international programs, or preparing for university applications.

Fifth, although this study used an integrated theoretical framework, it did not quantitatively test relationships among the key concepts. Future studies could develop mixed-methods or quantitative research designs to examine how autonomy, competence, relatedness, parental trust, teacher support, school–family communication, and academic pressure interact to predict adolescent wellbeing. Such studies could help identify which factors have the strongest protective or risk effects across different educational environments.

Future research may also examine the growing role of digital technologies in adolescent education and wellbeing. Online learning platforms, AI-assisted tutoring, social media, digital monitoring tools, and algorithm-based learning systems are increasingly shaping students' school and family experiences. These technologies may provide personalized support and learning flexibility, but they may also intensify comparison, surveillance, performance anxiety, and self-management pressure. Understanding how digital education interacts with parenting, schooling style, and adolescent mental wellbeing will be an important direction for future studies.

In conclusion, this study argues that Eastern-style and Western-style schooling should not be understood as simple opposites. Both educational systems can generate psychological resources as well as psychological risks. The key issue is not which model is inherently superior, but how schools and families jointly shape adolescents' experiences of pressure, support, autonomy, recognition, and self-worth. Adolescent mental wellbeing requires more than reducing academic stress. It also requires trust, emotional recognition, meaningful guidance, opportunities for rest, respect for boundaries, and space for students to develop identities and values beyond academic performance.

Works Cited

Baumrind, D. (1971). Current patterns of parental authority. Developmental Psychology, 4(1, Pt.2), 1–103. https://doi.org/10.1037/h0030372

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard university press.

Chao, R. K. (1994). Beyond parental control and authoritarian parenting style: Understanding Chinese parenting through the cultural notion of training. Child Development, 65(4), 1111–1119. https://doi.org/10.2307/1131308

Chao, R. K. (2000). The parenting of immigrant Chinese and European American mothers: Relations between parenting styles, socialization goals, and parental practices. Journal of Applied Developmental Psychology, 21(2), 233–248. https://doi.org/10.1016/S0193-3973(99)00037-4

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. Psychological Bulletin, 113(3), 487–496. https://doi.org/10.1037/0033-2909.113.3.487

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Hassan, A., Jamaludin, N. S., Sulaiman, T., & Baki, R. (2010). Western and Eastern educational philosophies. In 40th Philosophy of Education Society of Australasia conference, Murdoch University, Western Australia.

Iyer, R. B. (2015). Blending East and West for holistic education. Educational Research and Reviews, 10(3), 244-248. https://doi.org/10.5897/ERR2014.2015

Liu, K. (2023). Middle school students' mental unwellness and academic performance in China: The effects of parental involvement. Plos one, 18(11), e0294172. https://doi.org/10.1371/journal.pone.0294172

Liu, J., Huang, Q., Park, H.-J., & Qin, D. B. (2025). Beyond collectivism and individualism: Chinese international students' perceptions of contemporary child-centered parenting. Family Relations, 74(4), 1850–1870. **https://doi.org/10.1111/fare.13157**

Ryan, R. M., & Deci, E. L. (2024). Self-determination theory. In Encyclopedia of quality of life and well-being research (pp. 6229-6235). Cham: Springer International Publishing.

Ryff, C. D. (2013). Psychological well-being revisited: Advances in the science and practice of eudaimonia. Psychotherapy and psychosomatics, 83(1), 10-28. https://doi.org/10.1159/000353263

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. Journal of personality and social psychology, 69(4), 719. https://doi.org/10.1037/0022-3514.69.4.719

Yamamoto, Y., Li, J., & Bempechat, J. (2022). Reconceptualizing parental involvement: A sociocultural model explaining Chinese immigrant parents' school-based and home-based involvement. Educational Psychologist, 57(4), 267-280.

Yu, S., Chen, B., Levesque-Bristol, C., & Vansteenkiste, M. (2018). Chinese education is examined via the lens of self-determination. Educational Psychology Review, 30(1), 177-214.https://doi.org/10.1007/s10648-016-9395-x

Zhang, X., Liu, J., Sopper, C., Wang, J. H., Yau, Y., & Scott, J. C. (2024). What is supportive parenting? Perspectives from Chinese immigrant caregivers with low-income living in the greater Boston area. Journal of Cross-Cultural Psychology, 55(4), 386-408. https://doi.org/10.1177/00220221241245442

A Low-Cost, Multi-Functional System for Real-Time Equine Health Monitoring
By Qiaoyi Chen

Abstract

Horses' performance and well-being can be seriously affected by undetected health issues, yet existing monitoring systems are often too expensive or lack comprehensive function. This study proposes a low-cost, multi-functional system for real-time equine health monitoring applicable during training, in stables, and on pasture. The system design integrates an IoT-based architecture that combines physiological monitoring, such as heart rate and body temperature sensors embedded in a saddle pad, with a GPS-enabled halter for location tracking and an AI-powered camera for behavioral analysis. Data transmission and mobile app integration allow continuous observation and owner alerts. A proposed testing framework and outcome assessment are discussed to predict the system's potential performance and identify possible technical challenges. Overall, this study highlights the proposal of an affordable and versatile monitoring solution to improve early detection of equine health issues, support data-driven management in equine training and care, and most importantly, enhance animal welfare.

Introduction

"Harigu might have to be put down…"

My mother's voice shattered the silence on a summer afternoon last year. I felt my chest tighten, my breath hitching as the words settled in. My first horse, who had carried me through my earliest days as a rider and nuzzled into my palms for treats, was slipping away. Tears blurred my vision, but I didn't hesitate.

"Take me to the barn, please," I told my mother, my voice trembling.

As we sped down the road, I fumbled to call Dr. Xue, the veterinarian. His voice was grave: "It's laminitis," a cruel disease that exposes the bone inside the hoof. Eventually, the infection would spread, leaving only two impossible choices: amputation or euthanasia. My mind raced, searching for anything that could save her. But there was nothing I could do.

I often wonder whether Harigu's condition could have been detected earlier in the presence of a multi-functional and real-time monitoring system. Horses cannot tell us when they are in pain, and subtle changes in their health are easy to miss until it is too late. This experience inspired me to explore how continuous health monitoring technology could help owners identify potential problems before they become life-threatening.

After Harigu's passing, I realized that something had to change in how we monitor horses. I researched several online sources for the equestrian monitoring system, and it turned out that none of the systems seem to include multiple functions. Moreover, the high cost of existing systems limits their accessibility and widespread adoption. The NIGHTWATCH Smart Halter, for instance, which tracks a horse's heart rate, respiration, activity, and movement patterns and automatically alerts owners when it detects signs of distress, costs around $1000 in 2020,

with a subscription fee of 3 months for $99/month, 6 months for $79/month, and 12 Months for $49/month in 2022 (“Reviews Needed”; “NIGHTWATCH® Smart Halter”). However, the system is offered through a subscription-based model, which may increase the long-term cost of use for horse owners (“NIGHTWATCH Smart Halter | Equine Colic, Foaling & Distress Monitor”). Meanwhile, a standard horse halter costs around $200, which is broadly affordable for horse lovers, including teenage riders.

Similar to the integrated monitoring system designed by the agricultural equipment manufacturer John Deere, which allows for the remote monitoring of the location and productivity of different pieces of machinery, I investigated how to design a more comprehensive system of monitoring for horses. Most of them merely have a single or two functions brought up by a device. For example, EquiTrace focuses on tracking and managing horse health records by recording body temperature and providing high-temperature alerts (EquiTrace). Wearable devices such as the Equimetre monitor physiological data including heart rate and speed, making them useful for training and competition (Arioneo). Although these technologies provide valuable information, each addresses only a limited aspect of equine welfare. Few systems combine physiological monitoring, behavioral analysis, GPS tracking, and real-time health alerts into a single platform.

By combining several different devices and their functions, I hypothesized to create a comprehensive and affordable monitoring system that can be used when horses are in the stable, in the pasture, or working out in the ring. Such a system can enhance the detection rate of many common illnesses or diseases by monitoring horses’ behaviors and physical conditions from multiple dimensions and sending alerts to users.

My paper outlines how a low-cost equine monitoring system can provide accurate real-time monitoring of both behavior and vital signs, including body temperature, heart rate, gps, and intestinal motility, during exercise and routine activities. The proposed monitoring system aims to integrate multi-functional monitoring capabilities to ensure the real-time management of horse health and safety. It continuously measures the horse’s heart rate and body temperature using embedded sensors, providing accurate physiological data. Equipped with a built-in GPS module, the device enables real-time geolocation tracking through a mobile application, allowing owners to monitor the horse’s position at any time. When abnormal heart rate or body temperature levels are detected, the system automatically generates alerts and transmits them instantly to the app. Additionally, the device is designed to interconnect with IoT-enabled cameras installed in the stable. By applying an AI-based behavioral recognition model, the system can identify unusual patterns in the horse’s activity, such as prolonged resting or reduced feeding behavior. Upon detecting potential health risks, the system promptly sends notifications via the app, enabling timely intervention and improves animal welfare.

Hypothesis

The proposed monitoring system will enhance the early detection of equine illnesses such as laminitis and ileus by continuously tracking horses’ behavioral and physiological conditions

across multiple dimensions and automatically sending alerts to owners, trainers, barn staff, or veterinarians.

Materials and Methodology

Materials

Two custom printed circuit boards (PCBs) will be designed and fabricated for this study: one mounted on the horse's chest to collect physiological data and another integrated with a stable camera to monitor behavioral patterns. The circuit designs will be created using Multisim software with a double-sided PCB layout to minimize the overall dimensions of the boards. The primary PCB incorporated several key modules, including a wireless Wi-Fi module for real-time data transmission, a GPS module for geolocation tracking, an infrared temperature sensor (MLX90614) for non-contact body temperature measurement, a heart rate monitoring module (AD8232) for electrocardiographic signal acquisition ("Wi-Fi in IoT"). The MLX90614 sensor, which costs approximately US$10.02 per unit, was chosen because it enables non-contact temperature measurement, reducing discomfort for the horse while providing reliable surface temperature monitoring ("Digital Plug"; "MLX90614ESF-BCH-000-SP"). The AD8232 module was selected due to its ability to acquire electrocardiographic (ECG) signals with relatively low power consumption and its widespread use in wearable physiological monitoring systems, as it saves space and cost (*AD8232 Datasheet*). A single sensor costs approximately US$7.37 and is widely available ("AD8232 Heart Rate Monitor"). A Wi-Fi module will enable stable real-time communication with the IoT server, while the GPS module allows continuous location tracking for horses during turnout or transport ("Wi-Fi in IoT"). Together, these modules provide the hardware foundation for real-time monitoring while keeping the overall system compact and cost-effective.

To enable reliable heart rate measurement, conductive electrode patches will be integrated into a custom-designed girth strap to maintain stable skin contact while minimizing discomfort during movement. IoT-enabled cameras will be installed within the stable to continuously capture video data for AI-based behavioral monitoring. All hardware components will be enclosed in protective housings to improve durability and ensure stable operation under both stable and outdoor environmental conditions.

For system development, Multisim will be used to design and optimize the PCB layout, ensuring efficient integration of the sensing, communication, and processing modules while minimizing the overall board size. Data communication between the hardware components, the IoT server, and the mobile application will be implemented using the MQTT protocol because of its lightweight architecture and efficient publish-subscribe communication model, making it well suited for real-time IoT applications (Bhanushali). A deep learning–based behavioral recognition model will be deployed on the IoT server to analyze real-time video streams and classify horse behaviors. This dedicated mobile application, which costs a total of approximately US$550-650,

will provide users with continuous access to physiological data, GPS location, behavioral status, and system-generated health alerts.

Methodology

The system integrates physiological monitoring, geolocation tracking, and behavioral analysis within a unified Internet of Things (IoT) framework. The chest-mounted PCB continuously measures heart rate and body temperature, while the camera-integrated PCB processes video streams and transmits them to an IoT server. The server acts as the central data hub, enabling communication between the hardware modules and the application on the user's phone.

Real-time heart rate monitoring will be achieved using the AD8232 module, which captures electrocardiographic (ECG) signals from the electrode patches embedded in the girth strap. It was selected because it provides low-power ECG signal acquisition while remaining compact and suitable for a wearable physiological monitoring tool (Analog Devices). To enhance conductivity and improve signal quality, a small amount of water can be applied to the contact interface between the horse's skin and the strap (Aarts et al.). The captured ECG data will be preprocessed locally on the PCB before being transmitted to the IoT server via the integrated Wi-Fi module.

Body temperature will be measured using the MLX90614 non-contact infrared temperature sensor, mounted on the chest PCB with its probe slightly exposed through a protective enclosure. The sensor will be positioned slightly above the horse's skin surface, enabling accurate temperature measurement without causing discomfort ("Digital Plug & Play").

In the proposed system, horse behavior will be monitored using IoT-enabled cameras installed within the stable environment. Video data captured by these cameras will be transmitted to an IoT server, where a deep learning–based behavioral recognition model will be deployed for analysis. The model would be designed to classify observable behaviors such as resting and eating based on posture, movement frequency, and duration. These behaviors were selected because changes in these actions have been recognized as important behavioral indicators of equine welfare and may reflect discomfort or compromised health (Lesimple).

Potential health risks will be inferred when abnormal behavioral patterns are detected, such as prolonged inactivity and reduced feeding behavior within a time period. In such cases, the system will be configured to generate an automated alert and transmit it to the mobile application for user notification. Although this study does not include full-scale training or validation of the behavioral recognition model, the proposed framework outlines how AI-based video analysis could be integrated into the monitoring system to support early identification of conditions associated with discomfort or digestive distress.

All physiological and behavioral data will be transmitted to the IoT server using the MQTT protocol. The server handles data storage, AI-based video analytics, and integration with the mobile application. The application is designed to continuously analyze physiological and behavioral data from the monitoring system, while also sending instant notifications when

abnormal physiological or behavioral actions are detected. Integrating both physiological and behavioral data into a single interface enables users to interpret multiple health indicators simultaneously.

Normal reference ranges for heart rate, respiratory rate, and body temperature will be programmed into the app based on established veterinary standards. For instance, a resting horse typically maintains a heart rate between 28 and 44 beats per minute and a respiratory rate of 8 to 16 breaths per minute, while body temperature usually falls within 37.2°C to 38.3°C (Scott and Martin 2). During exercise, these values increase naturally depending on intensity. Heart rate may exceed 200 beats per minute during strenuous exercise, while body temperature also rises drastically, particularly in warm environmental conditions (Scott and Martin 4).

The app's algorithm compares real-time readings to these reference ranges and identifies deviations that may indicate potential health concerns. When heart rate fails to return toward normal within 20 minutes after exercise, or when body temperature exceeds 39.5°C, the system triggers a health alert ("Do You Know What Your Horse's Vital Signs Should Be?"). Similarly, an unusually low temperature below 36.5°C or significantly reduced gastrointestinal motility may signal hypothermia or digestive complications such as colic or impaction ("Horse Colic"). The app categorizes alerts by severity, ranging from "attention required" to "critical", and immediately sends notifications to the owner, trainer, barn staff, or veterinarian. This mechanism ensures timely intervention, enabling early detection of issues such as laminitis, ileus, or heat stress based on multi-dimensional data trends.

Testing Protocol

Tier 1: Laboratory Testing: The first phase will focus on verifying sensor accuracy and system reliability in a controlled environment. Five to ten healthy horses can be selected for baseline testing under standing or sedated conditions to minimize movement interference. Body temperature readings from the saddle pad's NTC thermistors will be compared against rectal thermometry, which is a method that uses a digital or mercury thermometer inserted 5–7 cm into the rectum for 30–60 seconds, to ensure the precision within ±0.5°C (Chen). Heart rate data collected via the girth's ECG electrodes will be cross-referenced with a Polar Equine chest strap, while intestinal motility measurements from the accelerometer will be validated against manual stethoscope auscultation (Chen). GPS functionality was assessed for static positioning accuracy, measuring error margins in meters. Additionally, motion artifacts will be simulated using a treadmill to evaluate sensor stability during trotting and cantering. Data transmission and app alert responsiveness will be recorded to confirm real-time monitoring feasibility.

Tier 2: Real-world Field Testing: In the proposed field-testing phase, up to twenty horses, including individuals with pre-existing conditions such as laminitis and ileus, would be equipped with the device over a three-month monitoring period. Vital signs (temperature, heart rate, intestinal motility and GPS location) will be continuously tracked during workouts, rest, and feeding. Behavioral anomalies (e.g., reduced intestinal sounds, elevated temperature) triggering AI alerts will be documented and compared against veterinary diagnoses. User feedback from

trainers and veterinarians will be further collected to assess false-positive rates. Environmental factors such as summer heat and winter cold will be incorporated to test seasonal adaptability, particularly for temperature calibration. Last but not least, due to the difference in testing environments, the system should be tested in different countries and in different climates (Chen).

Proposed Results

The outcome assessment focuses on evaluating the projected performance of the proposed equine health monitoring system under the outlined testing protocols. The expected results are based on how accurately and consistently the system can capture and interpret physiological and behavioral data compared with standard veterinary measurements. During simulated testing, the heart rate and temperature sensors are expected to maintain an accuracy within ±5% of values obtained from certified reference devices (Chen). The communication module should demonstrate stable wireless transmission over typical stable and paddock distances, ensuring continuous data flow with minimal latency.

The data analysis component of the mobile application is expected to correctly classify normal and abnormal readings according to preset thresholds and trigger appropriate alert levels with at least 95% detection consistency. Furthermore, the integrated AI-based behavioral monitoring system should be able to identify irregular activity patterns, such as signs of discomfort or reduced movement, with a projected recognition accuracy above 90% under adequate lighting and camera positioning (Chen). Usability evaluation anticipates that the application will remain intuitive and responsive, enabling non-specialist users to access real-time updates and health summaries easily. Collectively, these predicted outcomes suggest that the proposed system could provide a reliable and practical tool for continuous equine health surveillance across diverse environments.

Discussion

The proposed monitoring system supports the hypothesis that a low-cost, multi-functional platform could enhance the early detection of equine health issues through integrated physiological and behavioral monitoring. By combining heart rate, body temperature, intestinal motility estimation, GPS tracking, and AI-assisted behavioral analysis within a unified IoT framework, the system addresses key limitations of existing equine monitoring technologies, which often focus on a single parameter or function only during rest periods.

Compared with commercial products such as smart halters or temperature loggers, the proposed system is designed to operate continuously across multiple environments, including training sessions, stables, and pastures. This broader applicability is particularly important for detecting conditions such as laminitis, ileus, or heat stress, which may develop gradually and present subtle behavioral or physiological changes before acute symptoms appear. The inclusion of AI-based behavioral monitoring further distinguishes the system by allowing non-invasive observation of activity patterns that may otherwise go unnoticed by human caretakers.

Although the system has not yet undergone full experimental validation, the proposed

testing protocols and outcome assessment suggest that it has strong potential as an affordable and scalable solution for equine health management. If implemented and refined through future trials, this approach could improve, reduce emergency veterinary interventions and support data-driven training decisions for riders, trainers, and veterinarians.

Limitations

Despite its potential advantages, this study has several limitations that should be acknowledged. First, the system remains at the prototype design and proposal stage, and full-scale experimental validation has not yet been conducted. As a result, accuracy and alert sensitivity performance are based on existing written and comparable technologies rather than actual data.

Second, physiological measurements such as heart rate and body temperature may be affected by motion, sensor placement variability, and environmental conditions, particularly during high-intensity exercise (Xie et al.). While the system design includes strategies to mitigate these effects, further controlled testing would be required to ensure consistent performance across different horse breeds and body types.

Third, the behavioral monitoring component relies on camera positioning, lighting conditions, and the availability of sufficient training data for the AI model. In real-world stable environments, low visibility or individual behavioral differences may reduce classification accuracy (Dell et al.). Additionally, the current design does not account for privacy concerns related to continuous video monitoring, which would need to be addressed through data governance and user consent protocols.

Finally, the proposed system focuses on early detection and risk identification rather than definitive medical diagnosis. Alerts generated by the system are intended to support decision-making by owners and veterinarians, not to replace professional clinical evaluation. Future development would require field testing and veterinary collaboration to improve reliability.

Conclusion

This study proposes a low-cost, multi-functional system for real-time equine health monitoring that integrates physiological sensing, geolocation tracking, and AI-assisted behavioral analysis within a unified IoT architecture. By enabling continuous monitoring across training, stable, and pasture environments, the system aims to address critical gaps in current equine health management technologies.

The system includes heart rate, temperature, intestinal movement monitoring, GPS tracking, and in-stable monitoring via realtime camera. Such a comprehensive system provides horse owners all-around care for their loved ones, ensuring their health conditions are secured under different conditions. All the data is integrated to one single screen, so users can check the data and real-time monitor whenever and wherever they want. In addition, with the alert system and AI chat box, horse-lovers never need to worry about emergency conditions or surgeries. By

integrating multiple monitoring functions into a single affordable platform, the system maximizes functionality while minimizing overall cost, offering a practical and cost-effective solution for continuous equine health monitoring.

Although the system has not yet undergone full experimental validation, the proposed design and outcome assessment demonstrate its potential to support early detection of health issues such as laminitis, ileus, and heat stress. The emphasis on affordability and usability further highlights its relevance for a broad range of equine caretakers, from professional training facilities to individual riders.

Future work will focus on large-scale field testing and collaboration with veterinarians to refine behavioral models and alert algorithms. With continued development, this approach has the potential to contribute meaningfully to equine welfare.

Works Cited

Aarts, Rhana Mackie, et al. “Technologies for Equine Welfare and Performance Monitoring under Field Conditions – Where Do We Stand?” Equine Veterinary Journal, vol. 58, no. 4, 2026, pp. 872–88. Wiley Online Library, https://doi.org/10.1111/evj.70092.

AD8232 Datasheet and Product Info | Analog Devices. https://www.analog.com/en/products/ad8232.html#documentation. Accessed 1 July 2026.

“AD8232 Heart Rate Monitor Module Kit ECG Remote Health Monitor Sensor.” eBay, https://www.ebay.com/itm/404551499423?_ul=CN. Accessed 7 July 2026.

admin. “Analog Devices Introduces Low-Power, Single-Lead Heart Rate Monitor.” DAIC, 13 Aug. 2012, http://www.dicardiology.com/content/analog-devices-introduces-low-power-single-lead-heart-rate-monitor.

Bhanushali, Rahul. “MQTT for Real-Time Apps.” Medium, 2 July 2020, https://medium.com/@rahulbhanushali/mqtt-for-real-time-apps-6e8934b0f818.

Chen, Shine. Personal Interview. 29 August 2025.

Dell, Anthony I., et al. “Automated Image-Based Tracking and Its Application in Ecology.” Trends in Ecology & Evolution, vol. 29, no. 7, July 2014, pp. 417–28. DOI.org (Crossref), https://doi.org/10.1016/j.tree.2014.05.004.

“Digital Plug & Play Infrared Thermometer in a TO-Can.” Melexis, https://www.melexis.com/en/product/MLX90614/Digital-Plug-Play-Infrared-Thermometer-TO-Can. Accessed 1 July 2026.

“Do You Know What Your Horse’s Vital Signs Should Be? - KPP.” Kentucky Performance Products, 26 May 2020, https://kppusa.com/do-you-know-what-your-horses-vital-signs-should-be/.

“EquiTrace - The Home of Horse Health Technology.” EquiTrace, https://equitrace.app/. Accessed 30 June 2026.

“Horse Colic.” Wikipedia, 29 June 2026. Wikipedia, https://en.wikipedia.org/w/index.php?title=Horse_colic&oldid=1361712344.

Lesimple, Clémence. “Indicators of Horse Welfare: State-of-the-Art.” Animals, vol. 10, no. 2, Feb. 2020, p. 294. www.mdpi.com, https://doi.org/10.3390/ani10020294.

“MLX90614ESF-BCH-000-SP.” DigiKey Electronics, https://www.digikey.com/en/products/detail/melexis-technologies-nv/MLX90614ESF-BCH-000-SP/5414795. Accessed 7 July 2026.

“NIGHTWATCH® Smart Halter™ | Equine Colic, Foaling & Distress Monitor.” NIGHTWATCH, https://www.smarthalter.com/. Accessed 30 June 2026.

Racehorse Heart Rate Monitor | Equimetre Flat & Harness - Arioneo. https://training.arioneo.com/en/racehorse-heart-rate-monitor-equimetre-flat-harness/. Accessed 30 June 2026.

“Reviews Needed on Horse Sensors: Nightwatch, Trackener, Arioneo, Equestic, Polar, Ceefit, Hylofit - Off Course.” Chronicle Forums, 3 Dec. 2020,

https://forum.chronofhorse.com/t/reviews-needed-on-horse-sensors-nightwatch-trackener-arioneo-equestic-polar-ceefit-hylofit/753391/3.
Scott, Brett D., and Mike Martin. Understanding Vital Life Signs in Horses.
Xie, Hongwei, et al. “State-of-the-Art Wearable Sensors for Cardiovascular Health: A Review.” Npj Cardiovascular Health, vol. 2, Oct. 2025, p. 53. PubMed Central, https://doi.org/10.1038/s44325-025-00090-6.

Development and Biodegradability Assessment of Eco-Friendly Bioplastics Derived from Banana Peel Waste By Jiatong Xie

Abstract

Over the past several decades, the reliance on plastic has increased rapidly. Plastic production around the world rose from approximately two million tonnes in the 1950s to more than 450 million tonnes over the next 70 years (1). Creating a very obvious environmental crisis, resulting in persistent pollution. Not only does it affect our environment, but also human health. The primary reason why plastics don't degrade, is because of their composition on synthetic polymers. Conventional plastics are made out of strong carbon-carbon bonds that can not be identified by naturally microbial enzymes (2). The microorganism lacks a specific enzyme that is required to break down these complexes. This study investigates the production of biodegradable plastics made from nontoxic and safe materials: banana peels, vinegar and starch along with glycerin. The main objective is to evaluate the extraction of starch within banana peels. Banana peels are one of the highest-containing fruits that are rich in starch (3), a natural polymer made from numerous glucose units joined together, which can be processed and modified to make it exhibit characteristics like the conventional plastics we use. Bananas also contain high concentrations of ash (4), which can contribute to structural integrity and mechanical properties of the bioplastic. To evaluate the biodegradable plastic, in this study, we did a series of tests, including a weight bearing test to determine the strength, flexibility of the bio plastic and using a range of different concentrations of glycerin (plasticizer), along with soil burial degradation to test the rate of decomposition, to see if it leaves any harmful substances behind. The bioplastic samples were then compared against the conventional plastics found in supermarkets. The test results concluded that banana peel-derived bioplastic exhibits traits just like the conventional plastics. Although banana peel-derived plastics may liquify once in contact with water, it strongly suggests that it's a potential for a wide range of industrial production.

Introduction

Purpose

Plastic pollution is one of the most critical environmental problems facing the world right now. It causes harm to not only us, but also our ecosystem. Plastics can take up to decades or even centuries to fully decompose. This study explores the potential of banana peel-derived plastic samples that are biodegradable in just within weeks. Regular plastics made out from synthetic polymers forbid microbes to break them down. Banana peels however is a widely thrown waste that we take for granted, because of the high contaminants of starch and ash, we realised, it may be a possible solution to help us solve the crises we face right now. We are investigating whether the starch in the banana peels can be modified into daily plastics we use and decompose much faster and reduce our reliance on harmful synthetic materials. Our goal

was to show that waste materials can be a valuable resource in helping us create more environmentally friendly solutions.

Hypothesis

It is hypothesized that if starch components from banana peel are extracted and mixed with varying amounts of glycerin, the resulting bioplastic will show mechanical properties similar to conventional plastic, including flexibility and weight-bearing capacity, while decomposing more rapidly in soil and not emitting any harmful substances.

Background research

The background research explores different methods for producing different flexibility of biodegradable plastics from banana peels. We investigated various natural plasticizers including glycerin, nitric acid, honey and vegetable oils, and found out that some of the natural plasticizers can lead to brittleness over time. We also researched binding and strengthening agents such as starch. We found that these banana peels contain starch, cellulose and pectin, with starch being the key component for bioplastic production. When blended and heated with water, starch polymer chains unwind and form a continuous network, creating a strong bond. Glycerol acts as a plasticizer by forming hydrogen bonds with these starch chains, preventing them from packing too tightly and maintaining the flexibility of the bioplastic. An increase in glycerol concentration leads to an increase in film thickness, moisture content and solubility in water (5). Vinegar (acetic acid) also aids in the polymer chain connection, resulting in a smoother structure. Therefore, this research supported us that our choice of glycerol should be our primary plasticizer because of its established effectiveness in promoting flexibility in starch-based bio-plastic.

The following table shows the time taken for different objects to decompose table 1:

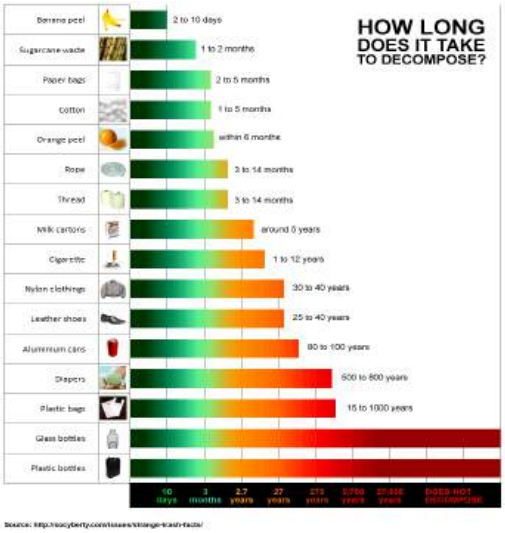


Table 1 (Hess)

Based on the data shown above from table 1, plastic bottles and plastic bags are one of the few things that take the longest to be fully decomposed. It takes around 15 to 1000 years for plastic bags and more than 200 centuries for plastic bottles, some may persist in the environment for extremely long periods and may not fully decompose under natural conditions. Even if the plastic has been fully decomposed, the plastic breaks down into toxic microplastics that enter the food chain, leach harmful chemicals into ecosystems, and threaten wildlife and human health (6), not only that, it leads to plastics accumulating in landfills and natural habitat, While they sit there, they clutter landfills, oceans, rivers, and soil instead of being broken down naturally. They can be eaten by animals or trap them due to them mistaking plastic for food, leading to their stomachs being blocked and their body being poisoned by chemicals and toxic materials. Most of the deaths to animals are caused by entanglement or starvation. Seals, whales, turtles, and other animals are strangled by abandoned fishing gear or discarded six-pack rings. Microplastics have been found in more than 100 aquatic species (7). As long-lasting waste slowly degrades, it can leak harmful chemicals into soil and water. This contaminates drinking water and damages plants and microorganisms that keep ecosystems healthy, damaging the whole environment and its source of health.

Whereas, banana peels are more environmentally friendly material because they naturally decompose in a shorter period of time compared to plastics. As shown in the table, it approximately takes up to 2 to 10 days to decompose. Banana peel also doesn't harm animals. As they come across banana waste, it won't poison them or trap them like plastic can, helping them to stay healthy and alive. Lastly, bananas disappear completely when they decompose leaving no microplastics behind them, whereas plastic breaks into tiny pieces that actually stay in the environment and enter food chains. When a banana peel decomposes, it can actually enrich the soil.

Components of banana peel table 2:

Table 1. Nutritional Component of Banana Peel (Essien et al. 2005; (Lee, Yeom, Ha, & Bae, 2010); Emaga et al. 2007).

Nutritional Component	Average Content, % DM
Starch	3.5 – 6.3
Resistant Starch	2.3 – 2.5
Dietary Fibre	47 – 53
Crude Fat	2.24 – 11.6
Crude Protein	5.5 – 7.87
Ash	9 – 11
Carbohydrate	59.51 – 76.58

Table 2 (Mohd et al.)

As shown in Table 2, banana peels contain high levels of starch, which serves as the primary matrix for bioplastic formation. During heating, starch forms a continuous polymer network that provides the main structure of the bioplastic (8). Having high percentage of ash can help make the plastic more durable, because it acts as fillers and can help increase the stiffness of the plastic. Proteins can act as natural binding agents, improving tensile strength and potentially participating in crosslinking to enhance structural integrity and water resistance.

Methodology

Materials

Banana peels	approximately 50 g
Distilled water	100 mL
Glycerin (glycerol)	5 ml (1 teaspoon)
Blender	1
Cloth filter or sieve	1
Beaker or pot	Capacity of at least 250 mL
Hot plate or stove	1
Stirring rod or spoon	1
starch	5 grams
Flat tray or mould	Approximately 15 cm × 15 cm
vinegar	4 grams

Procedure

1. Preparation of starch
   mix 5 grams of starch with 4 grams of vinegar
2. Preparation of Banana Peel Paste
   approximately 50 g were washed thoroughly and cut into small pieces. The pieces were placed into a blender with 100 mL of distilled water and blended for 1–2 minutes until a smooth paste was obtained.
3. Extraction of Banana Peel Mixture
   The blended paste was filtered through a cloth filter into a beaker. The filter was squeezed to extract as much liquid as possible. The filtrate was allowed to stand for 10–15 minutes before the clear liquid was carefully poured off, leaving approximately 80 mL of the thicker mixture.

4. Heating the Mixture
   The 80 ml banana peel mixture was heated at approximately 70–80 °C while being stirred continuously for 10–15 minutes.
5. Addition of Glycerin and starch solution
   Five millilitres (1 teaspoon) of glycerin were added to the heated mixture, along with the starch solution. The mixture was heated and stirred for an additional 5 minutes until it became thick, sticky, and slightly translucent.
6. Formation of the Bioplastic Film
   The mixture was poured onto a tray lined with aluminium foil and spread evenly to a thickness of approximately 2–3 mm.
7. Drying
   The bioplastic film was left to dry at room temperature (25–30 °C) for 48–72 hours. Once completely dry, it was carefully removed from the tray for observation.
8. Soil burial decomposition test
   Once completely dry, the bioplastic film was carefully removed from the tray. For the decomposition test, the film was buried in soil and observed daily for 6 consecutive days.



figure 1 (starch and vinegar)



figure 2 (boil banana)



Figure 3 (filter banana)



figure 4 (pour the solution and filtrate into the pan)



figure 5 (boil at low heat)



figure 6 (boil until it thickens)



figure 7 (spread it flat)



figure 8 (let it dry for 48 hours)

To test out the rate of decomposition, we buried two samples into the soil.
The following table shows the observed changes in the bioplastic during a six-day soil burial test:

Day	Observation
Day 1	Bioplastic remains firm with no visible changes.
Day 2	The surface becomes slightly softer after absorbing moisture.
Day 3	Edges begin to soften
Day 4	Material becomes more flexible and starts to lose its original shape.
Day 5	Signs of microbial starts showing up
Day 6	Bioplastic begins to break down into smaller pieces, indicating the start of decomposition.

Result and discussion

Initially, it was hypothesized that the starch naturally present in banana peels would be sufficient to solidify the bioplastic solution. However, the initial formulation, consisting of a 1:10 ratio of glycerin to banana peel extract, did not solidify even after 72 hours. This indicated that the starch content in the banana peel extract alone was insufficient to form a stable bioplastic. Consequently, additional starch was incorporated into the formulation, and different starch concentrations were tested to determine the optimal amount for bioplastic formation. Table 1

summarizes the starch concentrations tested and the resulting state of the bioplastic after 48 hours

Starch (grams)	State of bioplastic after 48 hours
0	Liquid
1	liquid
2	Liquid
3	Liquid but thicker
4	Solid but still a bit moist
5	solid

Our results confirmed that banana peels can indeed be processed to create a bioplastic material exhibiting characteristics similar to conventional plastics. The freshly prepared banana peel bioplastic initially exhibited a clear transparent appearance, however, over time, the material gradually changed to a brown, translucent colour. This colour change is due to the presence of Polyphenol oxidase (PPO) is one of the primary enzymes responsible for enzymatic browning in banana peels (9). Upon exposure to oxygen or physical damage, PPO catalyses the oxidation of phenolic compounds, including caffeoylquinic acid, catechins and dihydrochalcones, into highly reactive o-quinones through its cresolase (monophenolase) and catecholase activities (10). These o-quinones subsequently undergo oxidation and polymerisation reactions to form brown melanin pigments, which are responsible for the observed browning of the bioplastic (9, 11). The degree and rate of browning is influenced by environmental factors such as exposure to oxygen, light, temperature and moisture. Therefore, reducing exposure to these conditions may help minimise enzymatic browning and maintain the appearance of the bio plastic for a longer period (12). However, we observed limitations primarily related to the relatively short polymer chains and weaker intermolecular forces inherent in banana peel-derived starch. This resulted in a bioplastic with reduced strength and load-bearing capacity, restricting its use to lightweight applications. While layering the bioplastic films increased overall strength, this approach introduced another constraint: the material's susceptibility to water absorption. When exposed to moisture, the bioplastic softens, loses structural integrity, and experiences a significant reduction in strength. This sensitivity to water is likely due to the hydrophilic nature of starch and the disruption of intermolecular bonds by water molecules. Further investigation is needed to explore methods for enhancing the water resistance of banana peel bioplastics, such as incorporating hydrophobic additives or crosslinking agents. These modifications could potentially improve the material's overall performance and expand its range of applications.

Conclusion

This research explores the global crisis of plastic pollution by creating a sustainable alternative from banana peels. We successfully produced bioplastic using banana peel pulp mixed with glycerol as a plasticizer. As hypothesized, the banana peel bioplastic was more flexible than unmodified material and decomposed much faster than standard polyethylene plastic, aligning with our background research that starch and cellulose in banana peels are readily broken down by microorganisms. The bioplastic had clear initial properties, though we observed it turned brown over time due to natural oxidation, and it softened when exposed to water. Components like starch, ash, and protein all played key roles in its structure. While it's currently best suited for lightweight uses, our findings show banana peels can be transformed from waste into a valuable resource. This project proves that eco-friendly alternatives to conventional plastics are possible using everyday organic waste, offering a way to reduce harmful pollution and move towards more sustainable materials.

Works Cited

1. Ritchie, Hannah, et al. Plastic Production Over Time. Our World in Data, [relevant URL if available].
2. "How Long Does It Take for Plastic to Decompose?" Encyclopaedia Britannica, https://www.britannica.com/science/How-Long-Does-It-Take-for-Plastic-to-Decompose. Accessed 7 July 2026.
3. "Starch Content of Fruits." FitAudit, https://fitaudit.com/categories/frt/starch. Accessed 7 July 2026.
4. Mohd, H. A., et al. "Banana Peels as a Bioactive Ingredient and Its Potential Application in the Food Industry." Journal of Functional Foods, vol. 92, 2022, p. 105054, https://doi.org/10.1016/j.jff.2022.105054.
5. Sanyang, M. L., et al. "Effect of Glycerol Plasticizer Loading on the Physical, Mechanical, Thermal, and Barrier Properties of Arrowroot (Maranta arundinacea) Starch Biopolymers." Scientific Reports, vol. 11, no. 1, 2021, https://doi.org/10.1038/s41598-021-91908-6.
6. Hess, Randi. "You Can Make a Difference. Use Less Plastic." Medium, 29 Oct. 2019, https://medium.com/@randihess777/you-can-make-a-difference-use-less-plastic-95fe2537d555.
7. Parker, Laura. "The World's Plastic Pollution Crisis, Explained." National Geographic, 28 May 2025, https://www.nationalgeographic.com/environment/article/plastic-pollution.
8. Tarique, J., et al. "Effect of Glycerol Plasticizer Loading on the Physical, Mechanical, Thermal, and Barrier Properties of Arrowroot (Maranta arundinacea) Starch Biopolymers." Scientific Reports, vol. 11, no. 1, 2021, p. 13900, https://doi.org/10.1038/s41598-021-93094-y.
9. Tilley, Andrew, et al. "Enzymatic Browning: The Role of Substrates in Polyphenol Oxidase Mediated Browning." Current Research in Food Science, vol. 7, 2023, p. 100623, https://doi.org/10.1016/j.crfs.2023.100623.
10. Gooding, Paul S., et al. "Molecular Cloning and Characterisation of Banana Fruit Polyphenol Oxidase." Planta, vol. 213, no. 5, 2001, pp. 748–757, https://doi.org/10.1007/s004250100553.
11. Taranto, Francesca, et al. "Polyphenol Oxidases in Crops: Biochemical, Physiological and Genetic Aspects." International Journal of Molecular Sciences, vol. 18, no. 2, 2017, p. 377, https://doi.org/10.3390/ijms18020377.
12. Sarkar, Shampa, et al. "Preventing Enzymatic Browning of Freshly Cut Green Bananas through Immersion in Normal Water, Lemon Juice, and Coconut Water." Food Science & Nutrition, vol. 12, no. 9, 2024, pp. 6612–6626, https://doi.org/10.1002/fsn3.4284.

The Effects of Acute Noise Exposure on Female Phonotaxis and Male Calling Behavior in *Gryllodes sigillatus* By Aiden Noh

1. Introduction

1.1 Anthropogenic Noise: Description and Prevalence

Anthropogenic noise pollution has recently emerged as a major issue in the environment affecting animal communication and reproduction as well as the functioning of ecological communities (Shannon et al.). Anthropogenic noise pollution increased consistently in terms of the level of anthropogenic noise rising by 5 decibels each decade. Currently, more than 80% of terrestrial habitats are characterized by ambient levels exceeding 50 decibels of sound pressure level (SPL), which is significantly higher than natural levels (Shannon et al.; Gomes et al.). However, ambient levels in forests vary between 30 and 40 decibels, while in urban areas, it rises to 70-90 decibels of SPL.

1.2 Acoustic Communication and Reproductive Success in Crickets

Given the widespread rise in anthropogenic noise levels, crickets provide a strong invertebrate model for exploring the influence of soundscapes on communication. Acoustic communication is a significant aspect of the reproduction of crickets, where males produce calling signals, and females locate the calling signals through auditory cues (Bent et al., "Anthropogenic Noise Disrupts Mate Searching"). Using stridulation, the production of sound by rubbing the file and scraper, specialized wing structures in crickets, males produce species-specific calling songs that have dominant frequencies of 4–6 kHz, with peak amplitudes measured at 75–90 dB SPL at 10 cm distance (Imaizumi and Pollack; Montealegre-Z et al.). Female crickets detect the calling signals through tympanal organs, which are found on the forelegs of the female, with the highest sensitivity found at a frequency of 4-6 kHz, corresponding with the male calling frequency (Imaizumi and Pollack). The female cricket follows a directed phonotactic path to the male cricket after detecting the song (Hirtenlehner and Römer). Female crickets do not passively select males: they select males with higher amplitude calls, longer chirp durations, and faster pulse rates, all traits correlated with male body size and genetic quality (Bent et al., "Anthropogenic Noise Disrupts Mate Choice").

1.3 Noise as a Reproductive Stressor

Because crickets rely so heavily on acoustic signals for reproduction, changes to the acoustic environment can act as a powerful reproductive stressor. Many animals rely on acoustic signals for critical life functions such as mate location, predator detection, and territory defense; this establishes the rapid change in environmental acoustics as a growing ecological stressor due to its global scale and perpetuity over time (Francis and Barber). Noise pollution has had a particularly large impact on crickets such as *Gryllodes sigillatus* as they rely heavily on acoustic communication and signaling for attracting mates (Francis and Barber). Traffic noise occurs at

65–75 dB SPL, overlaps with cricket calling songs, and masks them, causing females to take 80% longer to locate males (Etzler et al.). This masking compromises females' abilities of detection (whether females hear the call) and discrimination (whether females distinguish call quality) (Bent et al., 2021; Etzler et al.). This impairment in phonotaxis reduces reproductive success, egg production, and offspring viability (Welsh et al.). Unlike vertebrates that can suppress background noise through neural gating mechanisms, crickets lack the neurobiological substrate for active noise suppression (Imaizumi and Pollack). Furthermore, males cannot indefinitely increase call intensity without incurring unsustainable energetic costs (Francis and Barber; Montealegre-Z et al.). Female crickets preferentially mate with males producing higher-amplitude, longer, and faster calls, suggesting that males experiencing reduced signal efficacy under noise would be under strong selection to increase call output or alter temporal structure (Gray). However, energetic and biomechanical constraints on call production may prevent such compensation, because acoustic signalling in *Gryllus* species is metabolically expensive and tightly linked to individual condition and morphology (Erreger et al.; Erreger and Schmidt). This creates a "double vulnerability," where females experience degraded signal perceptions and males are unable to compensate through song modifications.

1.4 Ecological and Applied Relevance

These individual-level disruptions in communication and reproduction have both ecological and applied consequences. Crickets have a crucial role in food webs as prey for birds, reptiles, mammals, and spiders. Impaired reproductive success due to anthropogenic noise affects these predatory groups and endangers biodiversity and ecosystem services (Francis and Barber). *Gryllodes sigillatus* is increasingly produced on an industrial scale for various usages and animal feed; hence, reproduction has become of importance to insect industries (Gomes et al.). Understanding these interactions has important applications in the prediction of decline and the conservation of animals raised in and for commerce (Gomes et al.).

1.5 Knowledge Gaps

Despite existing research, there are three significant gaps that have not been addressed in existing research. Firstly, there is no conclusive discrimination between whether phonotactic deficits result from acute remodeling of the auditory system or from reversible habituation to noise. Secondly, there is no data on whether male crickets modulate their calling song in terms of amplitude, frequency, and pulse rate following acute noise exposure. Lastly, there is an alarming lack of similar research on *Gryllodes sigillatus*, a species with a similar ecological role and an increasingly prominent function in entomological trade, in comparison to research on *Gryllus bimaculatus* and *Acheta domesticus,* species that have been extensively studied in numerous investigations on the effects of noise on development. While existing research on *Gryllus bimaculatus* has centered on anthropogenic noise, similar research on *G. sigillatus* has not been conducted, in spite of its ecological significance (Bent et al., "Anthropogenic Noise Disrupts Mate Searching"; Welsh et al.).

1.6 Present Study

To address these gaps, this study examined how acute exposure to anthropogenic noise affects female phonotaxis behavior and male calling behavior in *G. sigillatus*. Crickets were tested either in acute anthropogenic noise (70 dB SPL) or in a quiet control condition (30-40 dB SPL of ambient noise), and adult behavior was recorded in standardized test arenas. For females, the study quantified changes in phonotaxis latency, success rate, and path efficiency toward male calling songs under noise versus quiet. This study examined whether acute noise exposure affects call amplitude, frequency, or temporal structure, independent of body morphology. By addressing this issue from an acute perspective and adult behavior, the research provides species-specific information on how anthropogenic noise affects communication ability in *G. sigillatus* and has implications for the development of sustainable insect culture practices as well as broader assessments of how anthropogenic noise influences invertebrate acoustic communication (Francis and Barber; Gomes et al.).

2. Research Questions and Hypotheses

2.1 Research Question

How does acute exposure to anthropogenic noise affect female *Gryllodes sigillatus* phonotaxis performance in terms of detection latency, success rate, and path efficiency?

Does acute noise exposure alter male *Gryllodes sigillatus* calling song parameters (amplitude, frequency, pulse rate) independent of body morphology?

2.2 Hypotheses

Female phonotaxis

1. If crickets are exposed to acute noise (70 dB SPL) during testing, then females will exhibit significantly impaired phonotaxis behavior, including increased search latency, reduced phonotaxis success rate, and reduced search path efficiency compared to quiet controls.

Male calling behavior

2. If crickets are exposed to acute noise during testing, then males will not show significant compensatory increases in call amplitude or shifts in dominant frequency or pulse rate, after controlling for body morphology.

Derived prediction

If female phonotaxis is impaired and male acoustic signals remain unchanged under noise, then the combination will produce a reproductive communication barrier in which females frequently fail to locate calling males despite active searching, leading to reduced mating success.

3. Research Design and Method

3.1 Subjects

Mixed-sex adult *Gryllodes sigillatus* were obtained from Carolina Biological Supply Company and acclimated for 3 days in tanks (W: 25.4 cm, L: 50.8 cm, H: 35.6 cm) before testing (Welsh et al., 2023). Crickets were sexed by presence of stridulatory file for males and ovipositors for females and housed at low density (~50/bin) to minimize stress and aggression (Bent et al.). Females and sexually mature males (3–7 days post-final molt) were selected for behavioral trials to standardize reproductive status (Hirtenlehner and Römer). All experiments were conducted in the Plainview-Old Bethpage High School Research Laboratory under adult supervision and in accordance with school safety guidelines for invertebrate research.

3.2 Housing Conditions

Habitats were maintained at 27–32 ℃, measured at substrate level using a strip thermometer, using 43 W halogen bulbs positioned 12-16 inches above the substrate on a 14:10 light:dark cycle controlled using mechanical outlet timers. Relative humidity was maintained at 60-80% with 3-6 centimeters of vermiculite (Vigoro 100% Vermiculite). Vermiculite provides superior moisture retention, odor absorption, and friable texture that supports cricket locomotion and egg-laying while preventing anaerobic bacterial growth common in wet peat or coir substrates (The Critter Depot). Food and water were provided as necessary using Fluker's high calcium cricket diet and fresh fruits and vegetables in shallow 6.5 cm diameter × 1.5 cm height ceramic dishes and damp sponges (W: 6.5 cm, H: 1.2 cm, L: 3.4 cm) with daily replacement to prevent mold growth and bacterial contamination (Welsh et al.).

3.3 Experimental Design

To test the effects of acute anthropogenic noise on phonotaxis and calling behavior in adult *Gryllodes sigillatus*, a 2 × 2 between-subjects factorial design was used (Etzler et al.). The primary independent variable was acoustic context with two levels: quiet control (30-40 dB SPL ambient noise) and 70 dB SPL traffic-noise playback (Francis and Barber). Dependent variables were separated by sex: for females, detection latency (time in seconds from start of call playback to when the female's entire body exits the shelter (when all legs clear the shelter opening), measured using video timestamps (Hirtenlehner and Römer), phonotaxis success rate (proportion of trials in which the female reaches a 5 cm radius of the calling speaker), and path efficiency (ratio of straight-line distance to total path length); for males call amplitude (peak dB SPL at 10 cm), dominant frequency (kHz), and chirp pulse rate (pulses per second) (Bent et al., 2021; Arak).

3.4 Phonotaxis Trials (Females)

Phonotaxis trials were conducted in a rectangular arena constructed from a clear plastic storage bin (interior dimensions 48.26 cm × 37.15 cm × 36.51 cm) lined with paper, providing a uniform visual background and minimizing visual cues that could influence female orientation (Etzler et al.). A speaker broadcasting a male calling song (4-6 kHz) at ~75 dB SPL, measured at 10 cm from the speaker using a calibrated Decibel X sound level meter, was placed against the

center of one end of the arena, either in quiet or overlaid with 70 dB SPL traffic noise (Greenfield et al.). The female release point was located at the center of the arena, where a removable shelter was provided to allow the cricket to habituate for 5 minutes before the trial began (Bent et al., "Anthropogenic Noise Disrupts Mate Choice"). This arena size was selected to allow short range phonotactic orientation while requiring females to actively detect, localize, and walk toward the sound source. Previous studies have used arenas of .5 × 1.5 m or 1.5 × 1.5 m (Hirtenlehner and Romer). The current arena follows the same principle, but is scaled to the smaller experimental space while maintaining the key features of active orientation, directed movement, and measurable path metrics.

The traffic noise stimulus consisted of a looped recording of recorded urban noise (concentrated between 2-8 kHz, overlapping cricket calling frequencies) played via Audacity and delivered through the full-range speaker, calibrated to 70 dB SPL (Brumm and Todt). In the quiet condition, the same calling song was played without added traffic noise.

Behavioral responses were video recorded from above using a fixed camera. For each trial, 3 metrics were extracted. Latency from stimulus onset to the female crossing a predefined line 5 cm from the speaker, success (binary, if the female reaches the speaker within 300s), and path efficiency, calculated as the straight-line distance from start to speaker divided by the total path length traced frame-by-frame using ImageJ (Hedrick).

3.5 Calling Behavior Trials (Males)

Calling behavior trials were conducted in individual sound chambers constructed with a clear plastic storage bin (48 × 37 × 36 cm interior) (Gerhardt and Huber). Adult males were in these chambers for a 5 minute acclimation period under the same temperature and light cycle as their housing bins to encourage natural calling (Greenfield and Roizen). Each male was recorded for a 5-minute period under either quiet control or 70 dB SPL traffic noise played from a speaker mounted outside the chamber wall calibrated using a sound level meter (Bennet-Clark).

Acoustic recordings were made using a microphone (Logitech Blue Snowball iCE USB) positioned 10 cm from the male. For each male, three call parameters were extracted using Audacity (v3.4.2): call amplitude, the peak dB SPL of the calling song, dominant calling frequency, the spectral peak (Hz) of the male's calling song, extracted via Fast Fourier Transform analysis of audio recordings (Montealegre-Z et al.), and pulse rate, the number of distinct pulses (individual chirp units) produced in 1-second interval of calling, measured from audio waveform (Montealegre-Z et al.). After recording, males were briefly anesthetized by cooling and body mass was measured using a digital scale, allowing call parameters to be analyzed while controlling variation in morphology (Gerhardt).

4. Results

4.1 Male Signaling

A ranked independent samples t-test revealed no significant difference in dominant calling frequency between quiet and noise conditions ($p = 0.115$, Cohen's $d = 0.73$): mean dominant frequency was 5.99 kHz (median = 6.00 kHz) in quiet and 6.11 kHz (median - 6.20 kHz) under noise.

Another ranked independent samples t-test revealed no significant difference in chirp rate between conditions ($p = 0.0668$, Cohen's $d = 0.81$), with quiet-condition males averaging 3.79 chirps per second (median = 3.90) compared to 3.39 chirps per second (median 3.40) under noise exposure.

A ranked independent samples t-test revealed a significant difference in call amplitude between conditions ($p = 0.000310$, Cohen's $d = 1.85$): quiet condition males averaged 80.1 dB (median = 80.3 dB), compared to 76.0 dB (median 75.5 dB) under noise.

A Pearson correlation revealed a significant positive relationship between body mass and call amplitude ($r = 0.452$, $p = 0.0265$, $R^2 = 0.204$), accounting for roughly 20% of the variance in amplitude.

4.2 Female Phonotaxis

A ranked independent samples t-test revealed significantly lower path efficiency under noise than in quiet conditions ($p < 0.00001$, Cohen's $d = 2.57$): mean efficiency was 78.3% (median = 81.0%) in quiet and 55.3% (median = 54.0%) under noise.

A ranked independent samples t-test revealed significantly greater latency to reach the speaker under noise ($p < 0.00001$, Cohen's $d = 2.85$): mean latency was 28.4 s (median 25.8 s) in quiet, versus 48.5 s (median= 47.7 s) under noise.

A fisher's exact test revealed a significant difference in phonotactic success between conditions ($p = 0.0361$, χ^2 (1) = 6.17, Cramér's V = 0.51): 71.4% of females located the speaker in quiet conditions, compared to 28.6% under noise.

5. Discussion

This study examined how acute anthropogenic noise exposure affects acoustic signaling in male *Gryllodes sigillatus* and phonotactic behavior in females. Male spectral and temporal signaling parameters were largely unaltered by noise exposure, while call amplitude was significantly suppressed and female phonotactic performance was severely impaired across all three measures. These findings suggest that acute noise does not trigger compensatory vocal adjustments in this species, but instead degrades the acoustic channel in ways that disrupt reproductive communication.

5.1 Dominant Calling Frequency

The absence of a significant change in dominant calling frequency is consistent with the prediction that *G. sigillatus* would not alter output in response to acute noise. The resonant properties of the stridulatory anatomy impose strong physical limitations on frequency modulation, and spectral purity and carrier frequency are tied to wing morphology and

biomechanical tuning, suggesting that frequency is likely more constrained that amplitude in this species (Montealegre-Z et al.). Although the Lombard effect, the tendency for vocalizing animals to shift frequency upward in noise, has been documented in some vertebrate species, evidence for analogous spectral plasticity in insects remains limited (Bent et al.). The medium-large effect size (Cohen's d = 0.73) warrants caution, however; with a larger sample, a modest frequency shift may have reached significance.

5.2 Chirp Rate

No significant difference in chirp rate was detected between conditions, remaining consistent with the prediction that acute noise would not clearly alter temporal patterning. Chirp rate is a temporally encoded feature females use to identify conspecific mates (Imaizumi and Pollack, 1999), and the energetic costs of stridulation may limit a male's capacity to upregulate call rate under masking conditions (Erregger et al., 2017). The observed trend, where males under the quiet condition produce slightly higher chirp rates, is consistent with suppression rather than compensation, though the large effect size (Cohen's d = 0.81) suggests the sample may have been underpowered to detect a real difference.

5.3 Call Amplitude

The significant reduction in call amplitude under acute noise directly contradicts the hypothesis of no compensatory change. Rather than maintaining or increasing amplitude to overcome masking, as documented in some species exploiting reverberant calling sites (Erregger and Schmidt), males tested under noise called significantly more quietly, suggesting noise suppresses rather than stimulates signaling effort. The positive correlation between body mass and amplitude confirms that body size is a meaningful predictor of signal strength in this species, and accounting for this covariation is important when interpreting amplitude differences between conditions (Gray).

5.4 Female Path Efficiency and Latency

The large reductions in path efficiency (Cohen's d = 2.57) and near-doubling of latency (Cohen's d = 2.85) strongly support the hypothesis that acute noise impairs female mate location. The effect sizes indicate minimal distributional overlap between conditions, suggesting noise disrupted both initial detection of the male signal and sustained directional movement toward the source. The mechanism likely involves masking at the level of the female auditory system, consistent with the literature in which signal-to-noise ratio degradation is the primary pathway through which anthropogenic noise disrupts acoustic communication (Francis and Barber). Female cricket phonotaxis is highly sensitive to masking interference even when male calls remain physically detectable, suggesting that noise impairs active directional tracing rather than detection alone (Bent et al., "Anthropogenic Noise Disrupts Mate Searching").

5.5 Female Phonotactic Success

The reduction in phonotactic success from 71.4% in quiet conditions to 28.6% under noise most directly captures whether noise exposure translates into a failed mating opportunity rather than a less efficient one. Noise exposure has similarly been shown to impair mate choice and reduce selectivity in female crickets (Bent et al., "Anthropogenic Noise Disrupts Mate Choice"). If females in noise-polluted environments consistently fail to locate males, reduced encounter rates could plausibly limit reproductive output—though confirming population-level implications would require longer term study (Shannon et al.). All three measures of female behavior support the hypothesis, strengthening the interpretation that noise impairs the full phonotactic sequence rather than any single component in isolation.

These findings may also have practical implications for cricket rearing, where noise from ventilation systems or urban proximity could reduce mating success in commercial colonies (Bent et al., "Anthropogenic Noise Disrupts Mate Choice"). The results also suggest that anthropogenic noise may create sex-asymmetric communication disruptions in acoustic insects, with females bearing the primary phonotactic costs while male signaling parameters remain largely stable (Erreger and Schmidt).

Although these findings are consistent with the interpretation that acute anthropogenic noise degrades reproductive communication through acoustic masking, other mechanisms may also contribute (Francis and Barber). In females, poorer phonotaxis under noise could reflect not only reduced detection or localization of the male signal, but also other behavioral effects of the acoustic environment (Francis and Barber). In males, the observed decrease in call amplitude may represent short-term suppression of signaling effort under aversive conditions rather than a fixed lack of compensatory capacity (Shannon et al.). Accordingly, the present design supports the conclusion that acute noise disrupts behavioral performance, but it does not fully isolate whether that disruption arises from signal masking, broader behavioral effects, or an interaction between the two (Shannon et al.).

Another interpretation is that acute noise also alters how crickets behave in the testing environment (Francis and Barber; Shannon et al.). Previous literature suggests that anthropogenic noise can change insect behavior in ways that are not limited to immediate signal masking, and cricket studies have shown that noise exposure can increase decision-making time and alter baseline auditory activity, indicating broader effects on behavior and sensory processing (Etzler et al.; Shannon et al.). In this study, the reduced female phonotaxis and lower male call amplitude may reflect a combination of acoustic interference and short-term behavioral suppression rather than a single mechanism (Francis and Barber; Etzler et al.). Because the present design did not separately measure arousal, general locomotor activity, or responses to a non-overlapping control noise, the results cannot fully distinguish among these possibilities (Etzler et al.; Welsh et al.). These findings are relevant to behavioral ecology because they suggest that noise may alter mate-location behavior more strongly than male signal production, shifting the balance of communication in acoustic insects. They also matter for insect husbandry, since noise in captive rearing environments may reduce reproductive success even when adult

individuals remain healthy and active. More broadly, the study contributes to research on how anthropogenic noise affects sensory communication in invertebrates.

6. Limitations and Future Research

There are several potential limitations worth noting in relation to these results. First, because this study used an acute noise exposure condition at 70 dB SPL, it is not possible to determine whether the effects increase with sound pressure or whether a threshold intensity is required before phonotactic and signaling behavior are diminished (Etzler et al.). In addition, because traffic noise overlaps with part of the frequency range used by male crickets during courtship, the study cannot distinguish between acoustic masking caused by overlap and behavioral suppression caused by exposure to a loud auditory environment (Shannon et al.). Future research would be to include a control condition with similar sound pressure but a different dominant frequency range to help separate these possibilities (Shannon et al.).

Second, the male results showed that call amplitude was significantly affected by acute noise exposure, but the findings for frequency and chirp rate had moderately large effect sizes while being insignificant, suggesting that limited statistical power may have caused these results (Etzler et al.). Using a larger sample size would help determine whether chirp rate and frequency are truly unaffected by acute noise or whether the current study lacked the power to detect subtle changes (Shannon et al.). Future studies could also examine amplitude more carefully by controlling for factors that may influence sound production, such as body mass, age, and thorax condition (Montealegre-Z et al.).

Third, female phonotaxis was measured in terms of latency, success, and path efficiency; however, the study failed to account for hearing thresholds, the neurological basis for phonotaxis, and locomotor speed (Imaizumi and Pollack). Thus, deficit in performance could be caused by signal detection failure, sound source localization errors, or general behavior modifications that are independent of the process of sound detection (Francis and Barber). In order to address this limitation in the study, future research might consider evaluating the baseline speed of the animals in addition to comparing silent and noisy trials of phonotaxis tests.

Fourth, although the experimental arena provided a controlled environment, it simplified the conditions found in nature (Hirtenlehner and Römer). In natural habitats, crickets encounter substrate vibrations, vegetation, temperature variation, and competing acoustic stimuli, which may affect phonotactic behavior. Future studies could use larger environmental ecosystems or semi-natural enclosures to test whether similar noise effects appear under more ecologically realistic conditions (Welsh et al.).

Lastly, the present study focused on acute noise exposure and, therefore, cannot establish whether the results are temporary, permanent, or accumulative with time (Welsh et al.). It was established that prolonged developmental exposure to noise can alter auditory and behavioral responses of crickets, suggesting that prolonged exposure may generate varied results compared to acute noise exposure only (Etzler et al.). A promising step for further research would be to

conduct a longitudinal study comparing acute noise, developmental noise, and combination exposure through several generations (Etzler et al.).

7. Conclusion

Ultimately, this study shows that exposure to anthropogenic noise leads to a disruption of reproductive communication in the *Gryllodes sigillatus* by reducing the phonotaxis ability of the females and reducing the male call amplitude without having any significant impact on male frequency or chirps. This shows that there is an asymmetric communication barrier that is faced by the females as a result of difficulty experienced when locating the male compared to producing a response signal. Together, these results contribute a new understanding of how anthropogenic noise may alter mating communication in acoustic insects and suggest that reproductive disruption may occur even when male signaling is only partially affected (Shannon et al.; Francis and Barber).

Works Cited

Arak, A. "Sexual Selection by Male-Male Competition in Natterjack Toad Choruses." Nature, vol. 306, no. 5940, 1983, pp. 261-62, doi.org/10.1038/306261a0.

Bennet-Clark, H. C. "Resonances in Insect Sound-Producing Systems." Acoustic Signals of Insects, edited by R. N. Möstermer and W. H. Thorpe, Springer, 1989, pp. 25-48, doi.org/10.1242/jeb.202.23.3347.

Bent, A. M., et al. "Anthropogenic Noise Disrupts Mate Searching in Gryllus bimaculatus." Behavioral Ecology, vol. 29, no. 6, 2018, pp. 1271-77, doi.org/10.1093/beheco/ary126.

---. "Anthropogenic Noise Disrupts Mate Choice Behaviors in Female Gryllus bimaculatus." Behavioral Ecology, vol. 32, no. 2, 2021, pp. 201-10, doi.org/10.1093/beheco/araa124.

Brumm, H., and D. Todt. "Noise-Dependent Song Amplitude Regulation in a Passerine Bird." Animal Behaviour, vol. 63, no. 4, 2002, pp. 891-97, doi.org/10.1006/anbe.2001.1968.

Erregger, B., et al. "Cranking Up the Heat: Relationships Between Energetically Costly Song Features and the Increase in Thorax Temperature in Male Crickets and Katydids." Journal of Experimental Biology, vol. 220, no. 14, 2017, pp. 2635-44, doi.org/10.1242/jeb.155846.

Erregger, B., and A. K. D. Schmidt. "Anthropogenic Calling Sites Boost the Sound Amplitude of Advertisement Calls Produced by a Tropical Cricket." Animal Behaviour, vol. 142, 2018, pp. 31-38, doi.org/10.1016/j.anbehav.2018.05.015.

Etzler, E. A., et al. "Road Noise Exposure over Development Increases Baseline Auditory Activity and Decision-Making Time in Adult Crickets." Communications Biology, vol. 8, 2025, article 280, doi.org/10.1038/s42003-025-07643-6.

Francis, C. D., and J. R. Barber. "A Framework for Understanding Noise Impacts on Wildlife: An Urgent Conservation Priority." Frontiers in Ecology and the Environment, vol. 11, no. 6, 2013, pp. 305-13, doi.org/10.1890/120183.

Gerhardt, H. C. "Female Mate Choice in Treefrogs: Static and Dynamic Acoustic Criteria." Animal Behaviour, vol. 42, no. 4, 1991, pp. 615-35, doi.org/10.1016/S0003-3472(05)80245-3.

Gomes, L., et al. "Influence of Anthropogenic Sounds on Insect, Anuran and Bird Acoustic Signals: A Meta-Analysis." Frontiers in Ecology and Evolution, vol. 10, 2022, article 827440, doi.org/10.3389/fevo.2022.827440.

Gray, D. A. "Female House Crickets, Acheta domesticus, Prefer the Chirps of Large Males." Animal Behaviour, vol. 54, no. 6, 1997, pp. 1553-62, doi.org/10.1006/anbe.1997.0584.

Greenfield, M. D., and I. Roizen. "Katydid Synchronous Chorusing Is an Evolutionary Stable Strategy." Nature, vol. 364, no. 6435, 1993, pp. 618-20, doi.org/10.1038/364618a0.

Hedrick, T. L. "Software Techniques for Two- and Three-Dimensional Kinematic Measurements of Animal Movement." Journal of Experimental Biology, vol. 211, no. 8, 2008, pp. 1221-28, doi.org/10.1088/1748-3182/3/3/034001.

Hirtenlehner, S., and H. Römer. "Selective Phonotaxis of Female Crickets Under Natural Outdoor Conditions." Journal of Comparative Physiology A, vol. 200, 2014, pp. 239-50, doi.org/10.1007/s00359-014-0881-7.

Imaizumi, K., and G. S. Pollack. "Neural Coding of Sound Frequency by Cricket Auditory Receptors." Journal of Neuroscience, vol. 19, no. 4, 1999, pp. 1508-16, doi.org/10.1523/JNEUROSCI.19-04-01508.1999.

Montealegre-Z, F., et al. "Structural Biomechanics Determine Spectral Purity of Bush-Cricket Calls." Biology Letters, vol. 13, no. 11, 2017, article 20170573, doi.org/10.1098/rsbl.2017.0573.

Shannon, G., et al. "A Synthesis of Two Decades of Research Documenting the Effects of Noise on Wildlife." Biological Reviews, vol. 91, no. 4, 2016, pp. 982-1005, doi.org/10.1111/brv.12207.

The Critter Depot. "Here's the Top 10 Best Bedding Material for Crickets." The Critter Depot, 2025, www.thecritterdepot.com/blogs/news/material-to-increase-your-population.

Welsh, G. T., et al. "Consistent Traffic Noise Impacts Few Fitness-Related Traits in a Field Cricket." BMC Ecology and Evolution, vol. 23, 2023, article 78, doi.org/10.1186/s12862-023-02190-2.

Dealing with an Asteroid Strike: The Impacts of the COVID-19 Pandemic and 2023 WGA and SAG-AFTRA Labour Stoppages on Toronto's Film Industry By Leila Popatia

Abstract

This paper explores the profound impacts of the COVID-19 pandemic and the ensuing 2023 Writers Guild of America (WGA) and Screen Actors Guild–American Federation of Television and Radio Artists (SAG-AFTRA) labour stoppages on Toronto's film industry. As Canada's largest production hub, Toronto experienced substantial instability and uncertainty across its theatres, production companies, and film workforce as release and production schedules were significantly delayed, disrupting film supply at an unprecedented scale. This study uses a qualitative and quantitative analysis of secondary sources, including industry reports, financial data, and media coverage, to examine how these disruptions affected film production, exhibition, and distribution in Toronto. The study explains how lockdowns caused a dramatic reduction in production levels and theatre attendance, leaving many workers without income. Meanwhile, with theatres closed and studios experimenting with streaming and home viewing, audience behaviour and consumption habits shifted, contributing to perhaps permanent changes in theatrical exhibition. While production in the city has surpassed pre-pandemic levels, attendance levels at movie theatres remain below pre-pandemic levels, likely due to lingering film supply shortages and increased consumer acceptance of streaming options. These findings suggest that the future success of Toronto as a film hub depends on its ability to adapt to both evolving audience behaviour and the economic priorities of studios.

Keywords COVID-19, labour strikes, streaming, movie theatres, film exhibition, film industry, Toronto, Canada, film festivals, infrastructure, pandemic

Author Biography

Leila Popatia is a junior at Branksome Hall in Toronto, Canada. Her interest in film and economics motivated her to investigate the twin effects of the COVID-19 pandemic and the labour stoppages, in rapid succession, on the film industry in her home city. The unprecedented scale of the disruption combined with the acceleration of streaming business models during the period raised questions about the long-term economic contribution of the industry in Toronto, including the sustainability of theatrical exhibition in particular. Through this research, Leila sought to answer her own questions about the future of film and movie theatres, both of which she loves.

Introduction

I remember going to see Christopher Nolan's Tenet in Toronto with my father and sister soon after emerging from a pandemic lockdown in 2020, expecting the usual crowds and buzz. Instead, the theatre was almost empty. There were no lines at the concession, and we were able to choose from rows of vacant seats. An unsettling silence filled the air before the show. The

following year, I went to see Spider-Man: No Way Home for my 12th birthday in 2021. By then, the experience was much more familiar. The theatre was busy, the audience cheered when Andrew Garfield and Tobey Maguire appeared, and I remember thinking that this was the sense of community I had missed for the past two years. These experiences reflect the dramatic shifts in the film industry during the COVID-19 pandemic. Theatres closed, attendance dropped, and studio streaming priorities overtook film exhibition. Two years later, the 2023 labour stoppages occurred and moviegoing diminished once again. For anyone who loves movies, understanding how these disruptions changed the industry feels essential for preserving and predicting the future of the film industry. This paper examines how the COVID-19 pandemic and the 2023 labour strikes reshaped Toronto's film industry, particularly in terms of production, exhibition, and audience behaviour.

Literature Review

Toronto's Film Industry and Reliance on Hollywood

Over the past decade, there have been two major disruptions to the film and television industry: the COVID-19 pandemic and the 2023 WGA and SAG-AFTRA strikes. The consequences of these two events reshaped theatrical exhibition, production, labour rights, and streaming business models on a global scale. However, most current and available discussion and analysis of the period centres on Hollywood, leaving scholarly gaps in the review of other North American film hubs. This literature review aims to critically examine existing research to gain insight into the impacts of these two disruptions on Toronto specifically.

Because Hollywood plays a dominant role in film distribution around the world, exhibition markets outside the U.S. are directly tied to Hollywood production. Canadian film culture operates within a media environment heavily influenced by U.S. productions, with Hollywood being the main source of entertainment in Canadian theatres, representing over 90% of box office revenue in recent years (Tinic 162, 163; Telefilm Canada, "Review of Canadian Moviegoing" 2, 4). Moreover, foreign location-and-service productions generated approximately CAD $4.86 billion in Canada in 2019, representing about half of the country's total film and television production volume (Canadian Media Producers Association 4, 70; City of Toronto, Protecting a Valuable Asset 2). In 2019, film, television, and digital-media production investment in Toronto reached CAD $2.2 billion, with foreign productions representing 72% of the city's film and television production activity (City of Toronto, Protecting a Valuable Asset 2, 4). At the time, the film industry in Toronto employed more than 30,000 people, leading the City of Toronto to identify it as one of the city's key economic drivers (City of Toronto, "Toronto's Screen Industry 2019" 2). Canada remains the largest foreign production centre for U.S. film and television given its skilled crews, tax credits, proximity and similarities, location-breadth, and exchange rate advantages. In 2019, before these major disruptions, film production in Canada accounted for over 180,900 full-time equivalent jobs and $12.8 billion of GDP for the Canadian economy (MPA Canada). Consequently, any disruption in Hollywood directly affects Canadian

exhibitors and production, as evidenced by a near 20% drop in film and television production volume in 2023-24, the year of the labour stoppages, and a 22% decrease in production-supported jobs (Casemore; Fortmueller 3).

COVID-19 and Audience Behaviour

While scholars have examined the economic effects of the pandemic and labour strikes on film production and exhibition, fewer studies have focused on consumer preferences and viewing behaviour. Media researchers argue that watching films is an active experience, meaning that audiences choose what they watch based on different emotional and social needs (Ungvarsky). The Uses and Gratifications Theory suggests that people consume media for reasons including entertainment, escapism, relaxation, and emotional connection (Ungvarsky). In the context of movie theatres, this means cinemas are not only competing with streaming services based on convenience and content quality, but also on the type of experience they can offer.

This idea becomes especially important when examining the recovery of Toronto's film industry after the pandemic and the 2023 labour strikes. Scholars in New Cinema History argue that going to the movies is often a social and cultural experience, and that people remember the feeling of being in a theatre just as much as the film itself (Van de Vijver 130, 140). This helps explain why the closure of theatres during the pandemic represented more than just an economic loss. In a city like Toronto, where large theatre chains, independent cinemas, and film festivals are all significant parts of the city's cultural fabric, the loss of shared moviegoing also disrupted an important social experience and the associated learning and dialogue, whether in media or between individuals (Van de Vijver 130, 140).

Studies on risk perception also provide insight into why reopening theatres did not automatically lead to full exhibition recovery despite a probable desire on the part of consumers for community and cultural recovery. Risk perception suggests that people respond not only to actual danger, but also to how safe they feel ("Risk Perception"). During COVID-19, people began associating shared indoor public spaces like cinemas with their health concerns (Cipolletta et al. 4, 10). As a result, returning to movie theatres was not only a decision about entertainment, but also a question of comfort and safety. This helps explain why some moviegoers quickly returned to cinemas after restrictions were lifted, while others remained hesitant long after theatres were legally allowed to reopen.

Scholars broadly agree that the COVID-19 pandemic severely disrupted theatrical exhibition as the global box office fell by 72% in 2020, not only because of theatre closures and perceived health risks but also because the number of films released theatrically in the United States and Canada declined by 66%, from 987 in 2019 to 338 in 2020, leaving even reopened cinemas with fewer new films to screen (Adgate). These losses were largely because cinemas rely on in-person audiences and concession sales (Pahl; Adgate; Mondello, "5 Years Ago, Movie theatres Closed"). Government lockdowns also accelerated the already established growth of studio-owned streaming platforms and disrupted traditional theatrical distribution models further,

with many studios shortening theatrical windows or experimenting with the economics of direct-to-streaming releases (Sweney; Artavia).

At the same time, while risk aversion may explain the slow recovery in theatrical attendance in the first months following a lifting of pandemic restrictions, cinema viewership remains far below pre-pandemic levels even in 2025 and 2026 year-to-date (Cineplex Inc. 2019 Annual Report; Cineplex Inc., 2025 Annual Report 10; Cineplex Inc. 2026 First Quarter), suggesting that reduced theatrical film supply and increased streaming adoption are perhaps the primary factors restraining full recovery today, years after the pandemic's end.

The 2023 WGA and SAG-AFTRA Strikes

Almost immediately after the film industry began recovering from the pandemic, the 2023 WGA and SAG-AFTRA strikes hit (Dwivedi and Chhonkar 1, 5). Estimates place the economic toll in the U.S. between $5 billion and $6.5 billion, with $1.4 billion to $1.6 billion in wages lost in the Los Angeles area alone (Carras). Approximately 45,000 industry jobs were lost in the United States between May and October 2023, creating a multiplier effect on thousands of supported sectors (Weprin). Both strikes were a response to a shifting entertainment landscape, including the widespread adoption of streaming and its unfamiliar business models, as well as the looming threat of artificial intelligence following the public launch of OpenAI's ChatGPT in November 2022, prior to the strikes. Hundreds of productions were once again halted soon after the disruptions of the pandemic, causing significant delays, and in some cases wholesale cancellations of films and television series.

Although the strikes occurred in the U.S., their effects extended to Canada due to the country's reliance on Hollywood production and exhibition. Many large-scale productions filmed in Canada are run by American studios, meaning that the U.S. strikes slowed production throughout Canada's film industry (Mudhar).

Existing research shows that both the COVID-19 pandemic and the 2023 writers' and actors' strikes created major instability within the global film industry by disrupting film production and release schedules, while accelerating the use of streaming. While scholars have addressed these impacts in a variety of national contexts, Canada's experience tends to be overlooked, perhaps because of its proximity to, and reliance on the U.S. industry.

Methods

This study uses a qualitative and quantitative analysis of primary and secondary sources, including industry reports, financial data, and media coverage, to examine how the COVID-19 pandemic and 2023 writers' and actors' strikes affected film production, exhibition, and distribution in Toronto. Specifically, the study looks at data from before the pandemic (2019), during the pandemic (2020-2021), during the strikes (2023), and in recovery (2022-present). The study gathers primary economic data from industry reports, including the annual reports of Canada's largest theatre chain, Cineplex, the Canadian Media Producers Association (CMPA), the Alliance of Canadian Cinema, Television and Radio Artists (ACTRA), government reports,

including Ontario Creates and the City of Toronto. Data collection focused on production spending, employment levels, theatre attendance, and box-office revenues, which were compared across different periods to identify trends and outliers.

To contextualize these findings into a broader timeline and gain insight into the global context of these disruptions, news sources such as The Globe and Mail, The Toronto Star, CBC, CityNews, Forbes, The Los Angeles Times, and The Hollywood Reporter were consulted. These sources provided information on the timeline of policy changes, theatre closures, government support programs, and effects of these disruptions on local workers and communities. A limitation of this research is that many sources reported on provincial and national trends, rather than specifically focusing on Toronto. This research addresses this limitation by considering the effects on Toronto-specific institutions such as the Toronto International Film Festival and independent cinemas in the Greater Toronto Area (GTA). These sources provide specific insight at a city level, enabling a comparison between national, provincial, and local experiences.

Findings

The findings of the research demonstrate that the COVID-19 pandemic and the 2023 writers' and actors' strikes permanently altered Toronto's film industry, in mostly negative ways, though also for the better in some cases. The shocks seem to have structurally impaired theatrical exhibition attendance, accentuated labour instability, and exposed Toronto's reliance on Hollywood. At the same time, the rise of streaming, while perhaps most damaging to exhibition and the long-term economic income of creators, has had some offsetting positive impacts on Toronto's production volumes.

The following sections delve into the details of how these two disruptions affected exhibition, audience behaviour, labour, and production infrastructure.

At its simplest, the pandemic appears to have made consumers less used to going to theatres and more dependent on streaming. Later, the strikes made it harder for theatres to recover, even as consumers wanted to return, because fewer major new films were being released and promoted. At the same time, the strikes were connected to problems that had become highlighted during the pandemic, especially unstable work and frustration over streaming data and revenue and how it would be shared, including if the emerging impact of AI began to be felt in the creative process.

The Collapse of Theatrical and Festival Exhibition During COVID-19

The COVID-19 pandemic marked one of the most severe disruptions to the film industry in decades, particularly to theatrical exhibition. In March 2020, government lockdowns forced cinemas across Canada to close, eliminating all in-person audiences and both ticket and concession sales (Cineplex Inc., 2020 Annual Report 6). Cinemas in Toronto, Canada's largest city and its economic hub, remained subject to capacity limits for two years, until March 2022 (Cineplex Inc., "First Quarter 2022" 3).

After the most severe lockdowns in 2020, movie theatres began reopening at different times across countries and regions in late 2020 and throughout 2021 (Cineplex Inc. "Cineplex Theatres across British Columbia Reopen Tuesday" 1). However, reopenings were often limited by capacity, with masking restrictions in effect. Consumer hesitancy to enter populated indoor spaces appears to have been a factor in the slow post-pandemic recovery in theatre attendance, exacerbated by inconsistent health authority advice and a series of false starts in reopening (Cineplex Inc. "Cineplex Details Plans to Move Forward." 1; Chen et al. 6–9). In Canada, for example, provincial governments repeatedly closed and reopened cinemas, often adjusting capacity limits, concession policies, and health regulations (Friend, "Ontario Movie Theatres to Reopen").

It is difficult to mathematically attribute moviegoer reluctance to lingering pandemic concerns versus the rise of streaming versus the slow recovery of film supply. However, there is evidence that film supply played a considerable role. When Top Gun: Maverick opened in May 2022, it earned approximately $156 million in the United States and Canada over the four-day Memorial Day weekend, setting a record for the holiday, even though the domestic box office overall remained approximately 40% below its 2019 level at that point (Mondello, "Top Gun: Maverick"; Bomey). A similar surge followed the release of Avatar: The Way of Water in December 2022. The film earned approximately $401 million domestically during that month, while Cineplex's December box-office revenue reached $48 million, equivalent to 65% of its December 2019 result (Box Office Mojo; Boxoffice Staff). These temporary increases suggest that audiences were willing to return for highly anticipated releases, even though overall box-office revenues remained below pre-pandemic levels.

In 2020, Cineplex revenue fell by 74.9%, to $418 million, while theatre attendance declined by more than 80% compared to 2019 (Cineplex Inc., 2020 Annual Report 17). The company reported a total net loss of $629 million, relative to $29 million in profit in 2019, which shows the profound economic impacts of the lockdowns and how much theatres rely on in-person audiences (Cineplex Inc., 2020 Annual Report 17). Even with government aid through programs like the Canada Emergency Wage Subsidy program, which provided the company with $57 million to support employee wages, Cineplex struggled to offset the challenges of the pandemic (Cineplex Inc. "Cineplex Details Plans to Move Forward." 1; Cineplex Inc., 2020 Annual Report 6).

At the time, Cineplex was to be acquired by the European theatre company Cineworld for $2.8 billion CAD in order to create the largest cinema business in North America (The Canadian Press). In June 2020, however, Cineworld backed out of the deal by invoking a legal clause claiming that the pandemic had created a "material adverse effect" making the deal too expensive (The Canadian Press). Cineplex sued Cineworld for wrongful termination and was awarded CAD 1.24 billion (Maddaus). These funds might have helped Cineplex recover from the pandemic; however, they were never received as Cineworld filed for Chapter 11 bankruptcy in 2022 (Reuters).

Independent cinemas in Toronto, on the other hand, have significantly smaller financial margins and savings compared to Cineplex. This forced them to experiment with alternative revenue methods. The Fox, a cinema in The Beaches, for example, sold the naming rights to seats in its theatre (Pahl). Along with The Revue Cinema and The Royal Cinema, it held virtual screenings and sold takeout concession foods (Pahl). Similarly, the Paradise Theatre promoted its affiliated restaurant and bottle shop (Pahl). Due in part to these methods and strong community support, most independent cinemas in Toronto managed to stay open, even while earning a fraction of pre-pandemic revenue through 2023 (Pahl).

In addition to impacting movie theatres, the pandemic also affected the 2020 Toronto International Film Festival (TIFF), one of the world's largest and most influential publicly attended film festivals (Toronto International Film Festival). In 2019, the festival screened 245 films, but in 2020 that number dropped to 50 (Dalke et al.). This decrease was owed to a decline in global film production and post-production, but also to the fact that many studios did not want to release films ahead of a recovery in theatre audiences given the importance of exhibition in increasing film profiles, including for awards recognition (Dalke et al.). Instead of filling theatres with critics and press, the festival held drive-ins, limited in-person screenings, and open-air events (Dalke et al.).

The Rise of Streaming

One of the most notable long-term consequences of the pandemic was the accelerated growth of streaming platforms, which has contributed to reduced theatre attendance, perhaps for the long-term. Although streaming services like Netflix, Amazon Prime, and Disney+ existed before the pandemic, their subscriptions soared because of lockdowns (Korsh). This success prompted the creation of more competitors such as HBO Max, Paramount+, and Peacock (Korsh). In 2020, streaming subscriptions passed 1 billion worldwide for the first time, 20.7 million of which can be attributed to Canadians (Telefilm Canada, 2020-2021 Annual Report 37; Faughnder). By 2022, more than 85% of Canadians subscribed to at least one streaming platform (Korzinski).

This prompted studios to experiment with different methods of releasing films. In 2021, for example, WarnerMedia initiated "Project Popcorn" where they released all of their theatrical films in theatres and on HBO Max simultaneously (McClintock). The plan was designed to quickly boost streaming subscriptions while theatres faced low attendance numbers due to the pandemic (McClintock). Popular films such as Dune, The Suicide Squad, Godzilla vs. Kong, and Space Jam: A New Legacy were included in the experiment (McClintock). Disney made similar changes with films like Mulan being released straight on to Disney+ (Spangler).

The pandemic provided studios with the opportunity to experiment, thereby "training" audiences to stay home for exclusive cinema-quality content and to build new habits (Ariccio and Theodorou 2, 5). Once viewers became accustomed to the convenience of streaming and shorter waits between theatrical and digital release, movie theatres had to work harder to prove their value, in a period marked by lingering consumer risk perception and fewer films available

for release (Song 1-2). Recent studies have found that the big screen and the social aspect of cinemas remain strong motivations for viewership, yet those same studies also identify high ticket prices compared with the value offered by streaming, as well as the greater convenience of home viewing, as major barriers (Telefilm Canada, Canadian Movie Consumption 20–21, 31, 36, 39–40). Audiences may not have lost interest in movies, but they do appear to have changed the circumstances in which they are willing to pay to see them (Telefilm Canada, Canadian Movie Consumption 5–6, 38–39). This helps explain why the recovery of theatrical exhibition has been so uneven and prolonged, with some films faring extremely well and others struggling to find an audience.

The effects of this shift were also visible within Toronto's film institutions. At TIFF, streaming platforms became increasingly important buyers of films, while the festival expanded its programming to include more television series alongside traditional feature films. TIFF CEO Cameron Bailey argued that the rise of streaming had changed both how films were produced and how they reached audiences, requiring festivals and filmmakers to adapt to an increasingly blurred boundary between theatrical film and streaming content (Zucker and Rajagopal). Streaming therefore altered not only the distribution strategies of studios and festivals, but also the viewing habits and expectations of audiences.

Toronto's independent cinema scene illustrates how exhibitors have responded to this increased audience selectivity. Rather than attempting to compete with streaming services on convenience or volume of content, theatres such as the Revue, the Fox, and the Paradise have increasingly emphasized repertory films, themed screenings, and special events that make moviegoing a distinct social experience. These cinemas have attracted growing numbers of younger viewers by offering programming and communal experiences that cannot be replicated through home streaming (Atad). Their popularity suggests that streaming has not eliminated demand for theatrical exhibition in Toronto, but has instead increased the importance of curation, community, and a clear reason to leave home.

An interesting subsequent research question, not explored here, is the extent to which plentiful financing and low interest rates during the pandemic drove a kind of land-grab for streaming subscriptions and content spending that was not grounded in a sustainable and profitable business case benefiting all industry constituents. With theatres hobbled, studios were free to experiment for possible upside with theatres bearing the downside. For example, a more recent trend among studios, even those with captive streaming businesses, has been to return to theatrical-first releases for major productions, and to adopt longer release windows. Even Netflix has announced a 49-day theatrical release window for its upcoming reboot of its Narnia intellectual property (del Barco). The reversion towards pre-pandemic windows suggests the shortening of theatrical windows and the simultaneous release of cinema content at home during the pandemic did not produce desired economic results for studios and may have been a failed experiment and an anomaly related to the pandemic and the associated closure of cinemas and low film-financing costs.

Attendance nevertheless remains lower than 2019 levels. Cineplex's pre-pandemic attendance levels were around 66.3 million people, while 2024 was only 42.9 million, a year that saw 94 wide releases relative to 120 in 2019 (Cineplex Inc. 2024 Annual Report 9). The attendance reduction of 35% relative to the supply reduction of 22% suggests compelling film supply is part but not all of the reason, with the rise of streaming and changing consumer preferences likely playing a role.

In response to the decrease in attendance, Cineplex has increased its per-patron revenue through higher ticket prices and premium screening formats (Cineplex Inc. 2024 Annual Report 9). Since 2020, the major Hollywood studios have also decreased their theatrical windows. Pre-pandemic films were kept in theatres for 75-90 days, but today, most films are shown for no more than a month before they are released on streaming services (Friend, "Cineplex Agrees to Shrink Theatrical Window"). This major shift suggests that streaming did not just expand during the pandemic, but it may have permanently altered consumer habits, making it more difficult for theatrical exhibition to recover.

Labour Instability

Beyond exhibition and changes in audience behaviour, the pandemic also caused widespread job loss across Toronto's film industry, particularly for freelance and entry-level workers. Ontario's only normal production activity in 2020 occurred in January and February, before the lockdowns (Ontario Creates 6). The province had 176 domestic productions, 85 fewer than the 261 recorded in 2019, and production spending fell by 30.8%, with many productions being put on pause, or even cancelled altogether (Ontario Creates 5–6).

This was an especially steep drop since 2019 saw record highs for production in Ontario (ACTRA Toronto 1). Lockdowns prevented the hundreds of thousands of people that typically work on set from being employed. At the height of the pandemic, an estimated 45,000 workers, employed directly or indirectly in the film industry, were out of work each month in Ontario (Topp and Jennings). Freelance workers and those looking for entry-level jobs were particularly vulnerable during these times because they relied on short-term contracts and in-person positions, creating significant income instability (Topp and Jennings). This caused many people to exit the film industry altogether. However, some sectors such as animation and VFX were able to work remotely, maintaining 85% capacity (Topp and Jennings).

The Expansion of Production Infrastructure

On the positive side, despite the lockdowns, an increase in demand for streaming content led to the expansion of Toronto's production infrastructure through the pandemic. Pinewood Toronto Studios, for example, started building five additional stages, while Cinespace Studios added 50,000 square feet to its Kipling Campus (ACTRA Toronto). This growth has been largely driven by the increasing demand for streaming content, with studios like Netflix opening an office in Toronto (Schneller). In just five years, from 2016 to 2021, Ontario's square feet of stage space grew from 2.1 million to 3.7 million (Schneller).

The exact date the pandemic ended is subjective, and varies by city, and it's even harder to provide an end point for its impact on the film industry. Many film buffs, however, suggest that the recovery of theatrical exhibition can be marked by the release of Spider-Man: No Way Home in December 2021. This was the first film since 2019 to earn more than $1 billion worldwide, becoming the sixth-highest grossing film of all time (Rubin). Unfortunately, however, almost immediately after the industry began recovering from the pandemic, another disruption hit.

The 2023 Strikes and Toronto's Reliance on Hollywood

On May 2, 2023, the Writers Guild of America (WGA) went on strike, demanding higher pay, streaming residuals, and better job security, including protections against AI in scriptwriting, following the public launch of OpenAI's ChatGPT the prior November (Donchenko 5–6; OpenAI). As previously discussed, the COVID-19 pandemic accelerated the growth of streaming while existing industry compensation structures had not adapted to this change. In traditional television, writers and actors could earn residuals when their shows were rerun. Compensation from streaming services, however, was disconnected from viewership and long-term success, including because of streaming's more opaque performance reporting relative to traditional measures like box office revenue and ratings. When workers saw their earning potential plummet, it helped catalyze the strikes. The WGA strike lasted 148 days, ending on September 27 (Donchenko 23–25). On July 14, SAG-AFTRA, the Screen Actors Guild–American Federation of Television and Radio Artists, went on strike hoping to achieve similar protections against streaming and AI (Donchenko 5–6). This strike lasted 118 days, ending on November 9, marking the first dual strike since 1960, and halting almost all film production across the U.S. for several months (McCluskey).

An important point, however, is that the strikes were shaped by the pandemic. As studios became more dependent on streaming, workers experienced layoffs, shorter contracts, and unstable income. By 2023, writers and actors were not only responding to new concerns like AI, but also to the economic instability that had grown during the pandemic. The strikes, therefore, revealed how deeply COVID-19 had intensified problems in the streaming economy.

Although the strikes were primarily centred in the U.S., they had significant ripple effects on Toronto, as many of the city's productions are overseen by American studios and workers (Mudhar). In fact, around 70% of Toronto's film and television production is from international projects, most of which are American (Mudhar). As a result, the strikes caused a dramatic decline in production in the city, with Toronto's Film Commissioner reporting an 80% drop in the number of active productions in 2023 (Mudhar). This had serious consequences for local film workers, with IATSE Local 873, Toronto's largest film technicians' union, reporting that wages for its members were down 89% (Mudhar). Similarly, foreign production spending in Ontario fell from $1.95 billion in 2022 to just $886.6 million in 2023, revealing the extent to which Toronto relies on the U.S. for film production and revenue (Mirza). This shows how closely Toronto's film industry is tied to Hollywood. Even though Toronto has many studios, crews, and

production resources, much of its work still depends on American projects coming to the city. When those projects stopped, Toronto workers were affected too, even though the strikes were happening in the U.S.

It is important to mention that the impact of the strikes was uneven across Toronto's film industry. Some Canadian productions, such as Murdoch Mysteries, were still able to continue running because they were entirely Canadian and therefore not tied to American unions (Barghiel).

Disruptions to Exhibition and Film Festivals During the Strikes

The halt in Hollywood production during the strikes created a shortage of new films, significantly impacting theatrical exhibition in Toronto by limiting box office revenues and reducing theatre attendance (Deschamps). Unfortunately, these delays did not end when the strikes did. The release dates of many large films like Mission Impossible: The Final Reckoning and several Marvel blockbusters were pushed deep into 2025, and even early 2026 (Kelley). This is where the relationship between the strikes and the pandemic becomes most visible. Theatres were already trying to rebuild attendance after COVID-19 had changed viewing habits and lowered theatre attendance. The strikes then disrupted the supply and promotion of new releases just as cinemas needed large films to convince audiences to return.

The 2023 Toronto International Film Festival, like other festivals, was also affected by the strikes, as actors were generally prohibited from promoting their films (Weaver). Many of the American films that premiered at TIFF that year lacked celebrity appearances, interviews, and Q&As, features that bring audiences to theatres and give films attention for awards (Weaver). Without celebrity promotion many of these films went unrecognized and underperformed box office projections. For example, Amanda Seyfried did not attend the TIFF premiere for Seven Veils nor was she able to promote the film (Weaver). This may have contributed to the film's low box office performance, grossing only $157,313 worldwide (Box Office Mojo). At the same time, the absence of Hollywood stars made way for Canadian filmmakers, with films like Bria Mack Gets a Life and Hell of a Summer getting more visibility (Vlessing). This is significant for Toronto's film industry because TIFF is not only a place where films are screened, but also a major publicity platform. Celebrity appearances, interviews, and red carpets help smaller films gain popularity, especially during awards season. Without that attention, the festival lost some of the excitement and media coverage that usually helps films gain momentum. For Toronto, this meant the strikes not only affected production, but also the city's role as an important festival and exhibition hub.

Conclusion

While this paper has its limitations, it offers a clearer picture of how two overlapping disruptions–the COVID-19 pandemic and the 2023 Hollywood labour strikes–reshaped Toronto's film industry. This study is limited by its reliance on secondary sources and by the difficulty of isolating the effects of streaming from strictly pandemic-associated closures and

behavioural changes. It is also difficult in some cases to separate Toronto from the wider Ontario, Canadian, and American industries, in part because much of the available data tends to aggregate film and television together or fails to disaggregate impacts on Toronto specifically. In addition, as industry supply recovers, some long-term effects may not yet be fully visible.

Overall, these findings show that the COVID-19 pandemic and the 2023 labour strikes should not be seen as separate disruptions with separate consequences. The pandemic changed the psychology of public gathering, accelerated streaming, and destabilized film labour, while the strikes, informed by the pandemic itself, made it harder for theatres to recover by delaying releases, reducing film promotion, and revealing how much Toronto depends on Hollywood. At the same time, the correlation between increasing film supply and increasing theatre attendance, as well as the return of theatrical windows provides cause for optimism, suggesting that the role of cinema remains intact, both as community experience but also as the economic "locomotive" that powers the launch of a film and determines its subsequent success in different viewing formats.

These findings raise concerns about Toronto's vulnerability, but also offer reasons for optimism about its future. The city's strength comes from more than production spending or studio space. It comes from the workers who make films possible, the theatres that give those films a public life, and the audiences that continue to gather. One lesson of the pandemic and the work stoppages is that Toronto is served well by diversity: a balance of domestic and U.S. productions, the participation of independent cinemas and multiplexes, and a vibrant festival community.

Works Cited

ACTRA Toronto. Operating Plan Progress Report 2020. 2021, https://www.actratoronto.com/wp-content/uploads/2021/03/2020-in-Review_020321.pdf.

Adgate, Brad. "The Impact COVID-19 Had on the Entertainment Industry in 2020." Forbes, 13 Apr. 2021, https://www.forbes.com/sites/bradadgate/2021/04/13/the-impact-covid-19-had-on-the-entertainment-industry-in-2020/.

Ariccio, Silvia, and Annalisa Theodorou. "Stay at Home behaviour during COVID-19: The Role of Person-Home Relationships." Journal of Environmental Psychology, vol. 97, Elsevier BV, Aug. 2024, article 102334–34, https://doi.org/10.1016/j.jenvp.2024.102334.

Artavia, David. "They Haven't Gone to the Movies since before COVID. Five Years Later, They're Still Choosing Streaming." Yahoo Entertainment, 13 Mar. 2025, https://www.yahoo.com/entertainment/they-havent-gone-to-the-movies-since-before-covid-five-years-later-theyre-still-choosing-streaming-110059978.html.

Atad, Corey. "Toronto's Indie Cinema Scene Is Booming: 'It's a Different Experience Than, Like, the Cineplex.'" Toronto Star, 25 Nov. 2023, https://www.thestar.com/entertainment/movies/toronto-s-indie-cinema-scene-is-booming-it-s-a-different-experience-than-like-the/article_7ea2e81f-42c1-56ea-99c1-2e59400bde59.html.

Barghiel, Naomi. "Why the Hollywood Strike Is Already 'a Big Deal' for Canada's Film Industry." Global News, 19 July 2023, globalnews.ca/news/9843272/hollywood-strike-canada-impact/.

Bomey, Nathan. "Movie Theaters Strike Back: 'Top Gun: Maverick' Delivers." Axios, 31 May 2022, https://www.axios.com/2022/05/31/top-gun-maverick-movie-theaters-amc.

Box Office Mojo. "Domestic Box Office for December 2022 - Box Office Mojo." Box Office Mojo, 2022, https://www.boxofficemojo.com/month/december/2022/.

Box Office Mojo. "Seven Veils (2023)." Box Office Mojo, IMDb.com, https://www.boxofficemojo.com/title/tt26671415/credits/.

Boxoffice Staff. "Cineplex Inc. Reports December 2022 Box Office Results: $48M, 65% of December 2019 - Boxoffice Pro." Boxoffice Pro, 13 Jan. 2023, https://www.boxofficepro.com/cineplex-reports-december-2022-box-office-results/.

Canadian Media Producers Association. Profile 2024: An Economic Report on the Screen-Based Media Production Industry in Canada. 2024, https://cmpa.ca/wp-content/uploads/2025/02/Profile-2024-ENG-Final.pdf.

Carras, Christi. "Here's How Many Jobs L.A. Lost during the Hollywood Strikes." Los Angeles Times, 7 Dec. 2023, https://www.latimes.com/entertainment-arts/business/story/2023-12-07/actors-writers-strike-economic-impact-report.

Casemore, Jamie. "Production Volume Down 18.5% in 2023–24, CMPA Reports." Playback, 19 Dec. 2024,

https://playbackonline.ca/2024/12/19/production-volume-down-18-5-in-2023-24-cmpa-reports/.
Chen, Y., et al. “The Road to Recovery: Sensing Public Opinion towards Reopening Measures with Social Media Data in Post-Lockdown Cities.” Cities, vol. 132, 2023, article 104054, https://doi.org/10.1016/j.cities.2022.104054.
Cineplex Inc. “Cineplex Details Plans to Move Forward.” Cineplex, 15 June 2020, https://mediafiles.cineplex.com/press-releases/Cineplex%20Moving%20Forward%20Press%20Relase%20June%2015%20FINAL_20200615120241_0.pdf.
Cineplex Inc. “Cineplex Theatres across British Columbia Reopen Tuesday.” Cineplex, 9 June 2021, https://mediafiles.cineplex.com/press-releases/06%2015%202021%20BC%20Cineplex%20Reopen%20Media%20Advisory_FINAL_20210618155736_0.pdf.
Cineplex Inc. First Quarter. 2022, https://mediafiles.cineplex.com/investor-relations/reports/Q1%202022%20Report.pdf
Cineplex Inc. First Quarter. 2026, mediafiles.cineplex.com/investor-relations/reports/Cineplex%20Q1%202026%20Report.pdf..
Cineplex Inc. 2019 Annual Report. 2019, https://mediafiles.cineplex.com/investor-relations/reports/Annual%20Report%202019%20FINAL.pdf
Cineplex Inc. 2020 Annual Report. 2020, https://mediafiles.cineplex.com/investor-relations/reports/Cineplex_AR_2020_ONLINE.pdf.
Cineplex Inc. 2024 Annual Report. 2024, https://mediafiles.cineplex.com/investor-relations/reports/Cineplex_2024_Annual_Report-Final.pdf.
Cineplex Inc. 2025 Annual Report. 2025, https://mediafiles.cineplex.com/investor-relations/reports/2025%20Annual%20Report.pdf.
Cipolletta, S., et al. “Risk Perception towards COVID-19: A Systematic Review and Qualitative Synthesis.” International Journal of Environmental Research and Public Health, vol. 19, no. 8, 2022, article 4649, https://doi.org/10.3390/ijerph19084649.
City of Toronto. Toronto’s Film Industry: Protecting a Valuable Asset Prepared by Marguerite Pigott, Film Commissioner and Director of Entertainment Industries Economic Development and Culture. 2022, www.toronto.ca/legdocs/mmis/2023/ec/bgrd/backgroundfile-239385.pdf.
City of Toronto. Toronto’s Screen Industry 2019: A Record Year. 2019, https://www.toronto.ca/wp-content/uploads/2021/02/9655-SCREENINDUSTRYTORONTO2019.pdf.

Dalke, A., et al. “How Has COVID-19 Affected the ‘T’ in ‘TIFF’?” Local Film Cultures: Toronto, 24 Dec. 2020, https://localfilmculturestoronto.wordpress.com/how-has-covid-19-affected-the-t-in-tiff/.

del Barco, Mandalit. “Netflix Goes to the Movies.” Texas Public Radio, 13 May 2026, https://www.tpr.org/arts-culture/2026-05-13/netflix-goes-to-the-movies.

Deschamps, Tara. “Cineplex CEO Sees ‘Pivotal Change’ after Hollywood Strikes Trigger a Q2 Loss.” CityNews Toronto, 9 Aug. 2024, https://toronto.citynews.ca/2024/08/09/cineplex-reports-21-4m-q2-loss-as-attendance-fell-from-year-ago/.

Donchenko, Nataliya. “What Factors Influenced the WGA, SAG-AFTRA Strikes against the AMPTP in 2023 and What Were the Missed Opportunities towards Reaching an Agreement?” Columbia University, 2024. Academic Commons, https://doi.org/10.7916/hfzn-kd25.

Dwivedi, Manish Kumar, and Aman Chhonkar. “Writers Guild of America Strike 2023.” Emerging Economies Cases Journal, 15 Dec. 2025, https://doi.org/10.1177/25166042251397697.

Faughnder, Ryan. “Streaming Milestone: Global Subscriptions Passed 1 Billion Last Year.” Los Angeles Times, 18 Mar. 2021, www.latimes.com/entertainment-arts/business/story/2021-03-18/streaming-milestone-global-subscriptions-passed-1-billion-last-year-mpa-theme-report.

Fortmueller, Kate. Hollywood Shutdown: Production, Distribution, and Exhibition in the Time of COVID. University of Texas Press, 2022.

Friend, David. “Cineplex Agrees to Shrink Theatrical Window to 17 Days for Universal Movies.” CBC News, 20 Nov. 2020, https://www.cbc.ca/news/entertainment/cineplex-theatrical-window-universal-1.5809862.

Friend, David. “Ontario Movie Theatres to Reopen Next Week.” CBC News, 10 July 2021, https://www.cbc.ca/news/canada/toronto/ontario-movie-theatres-1.6097767.

Kelley, Sonaiya. “All the Major Movies and TV Shows Delayed by the Strikes.” Los Angeles Times, 19 Sept. 2023, https://www.latimes.com/entertainment-arts/business/story/2023-09-19/tv-shows-movies-delayed-list-writers-strike-sag-aftra.

Korsh, S. “COVID-19 and Its Impact on Television Streaming Services.” Michigan Journal of Economics, 26 Sept. 2020, https://sites.lsa.umich.edu/mje/2020/09/26/covid-19-and-its-impact-on-television-streaming-services/.

Korzinski, David. “Streamed Out: Cost of Living Crunch Prompts One-In-Three to Cancel Streaming Subscriptions in Last Six Months.” Angus Reid Institute, 27 Oct. 2022, angusreid.org/netflix-streaming-canada-cord-cutting-tv-landlines/.

Maddaus, Gene. “Cineplex Wins CA$1.2 Billion Judgment in Busted Cineworld Sale.” Variety, 15 Dec. 2021, https://variety.com/2021/film/news/cineplex-cineworld-sale-suit-1235133657/.

McClintock, Pamela. “‘Project Popcorn’: WarnerMedia’s Box Office–HBO Max Experiment Gets Mixed Results.” The Hollywood Reporter, 27 Oct. 2021, https://www.hollywoodreporter.com/business/business-news/warnermedias-project-popcorn-box-office-hbo-max-experiment-gets-mixed-results-1235037096/.

McCluskey, Megan. “What Happened When SAG and the WGA Went on Strike in 1960.” Time, 14 July 2023, https://time.com/6294777/sag-wga-strike-1960/.

Mirza, T. “Canada’s Service Sector Still in Catch-Up Mode Post-Strikes.” Playback, 2 July 2024, https://playbackonline.ca/2024/07/02/canadas-service-sector-still-in-catch-up-mode-post-strikes/.

Mondello, Bob. “5 Years Ago, Movie theatres Closed. NPR’s Movie Critic Looks Back at COVID-19.” NPR, 9 Mar. 2025, https://www.npr.org/2025/03/09/nx-s1-5320161/covid-19-movies-hollywood-movie-theatres.

Mondello, Bob. “‘Top Gun: Maverick’ Boasted a Record Opening Weekend. Where’d the Numbers Come From?” NPR, 31 May 2022, https://www.npr.org/2022/05/31/1102260593/top-gun-maverick-boasted-a-record-opening-weekend-whered-the-numbers-come-from.

MPA Canada. “Profile 2019: Economic Report on the Screen-Based Media Production Industry in Canada.” Mpa-Canada.Org, 2021, https://www.mpa-canada.org/research_docs/profile-2019-economic-report-on-the-screen-based-media-production-industry-in-canada/.

Mudhar, Raju. “How the Hollywood Writers’ Strike Is Impacting Toronto.” Toronto Star, 22 Sept. 2023, https://www.thestar.com/news/gta/how-the-hollywood-writers-strike-is-impacting-toronto/article_0f4f83ff-5ac5-5c4b-bad9-e3a9bdef73e0.html.

Ontario Creates. Annual Report 2020–2021. 2022, https://www.ontariocreates.ca/uploads/About-Us/Eng/Annual-Report/OC_AR21_EN_V7_AODA_FINAL_Tabled-Jan-25-2022.pdf.

OpenAI. “Introducing ChatGPT.” OpenAI, 30 Nov. 2022, openai.com/index/chatgpt/.

Pahl, J. “After a Year of Lockdowns, Toronto Movie Theatres Rely on Creative Solutions to Stay Afloat.” The Varsity, 14 Feb. 2021, https://thevarsity.ca/2021/02/14/after-a-year-of-lockdowns-toronto-movie-theatres-rely-on-creative-solutions-to-stay-afloat/.

Reuters. “UK Movie Chain Cineworld Files for Chapter 11 Bankruptcy in U.S.” CBC News, 7 Sept. 2022, https://www.cbc.ca/news/business/cineworld-bankrupt-1.6574698.

“Risk Perception.” ScienceDirect Topics, Elsevier, https://www.sciencedirect.com/topics/earth-and-planetary-sciences/risk-perception.

Rubin, Rebecca. “Spider-Man: No Way Home Becomes First Pandemic-Era Movie to Smash $1 Billion Milestone Globally.” Variety, 26 Dec. 2021, https://variety.com/2021/film/box-office/spiderman-billion-dollars-box-office-pandemic-1235143308/.

Schneller, Johanna. “How Streaming Turned Toronto into Studio City Extraordinaire.” The Globe and Mail, 4 Feb. 2022, https://www.theglobeandmail.com/arts/article-how-streaming-turned-toronto-into-studio-city-extraordinaire/.

Song, J. “The Evolution and Impact of Streaming Services: Changing the Media Landscape.” Global Media Journal, vol. 22, no. 72, 2024, pp. 1–3, https://www.globalmediajournal.com/open-access/the-evolution-and-impact-of-streaming-services-changing-the-media-landscape.php?aid=95170.

Spangler, Todd. “Mulan Available to All Disney+ Subscribers in December, No Extra Fee.” Variety, 2 Sept. 2020, https://variety.com/2020/digital/news/mulan-disney-plus-all-subscribers-no-extra-charge-1234757666/.

Sweney, Mark. “Studios Bypass Cinemas with Lucrative Lockdown Premieres.” The Guardian, 2 May 2020, https://www.theguardian.com/film/2020/may/02/studios-bypass-cinemas-with-lucrative-streaming-premieres-the-end.

Telefilm Canada. A Mobilizing Effect 2020-2021 Annual Report. 2021, telefilm.ca/wp-content/uploads/telefilmar21enfinal.pdf.

Telefilm Canada. Canadian Movie Consumption: Exploring the Health of Feature Film in Canada. ERm Research, 2023, https://telefilm.ca/wp-content/uploads/2024/02/2023-Canadian-Movie-Consumption-Exploring-the-Health-of-Feature-Film-in-Canada.pdf.

Telefilm Canada. Review of Canadian Moviegoing and Distribution: Year 2025. Strategic Development and Business Intelligence Department, Feb. 2026, https://telefilm.ca/wp-content/uploads/2026/02/TFC-B-O-Report-2025-ENG.pdf.

Tinic, S. A. On Location: Canada’s Television Industry in a Global Market. University of Toronto Press, 2005.

The Canadian Press. “Cineworld Scraps $2.8-Billion Takeover of Canada’s Cineplex, Setting Up Legal Battle.” CBC News, 13 June 2020, https://www.cbc.ca/news/business/cineworld-takeover-cineplex-termination-1.5573054.

Topp, Brent, and Christina Jennings. Film and Television in the Age of COVID-19, and After. Public Policy Forum, 15 Oct. 2020, https://ppforum.ca/publications/film-and-television-in-the-age-of-covid-19-and-after/.

Toronto International Film Festival. “Our Mission.” TIFF, https://www.tiff.net/mission.

Ungvarsky, Janine. “Uses and Gratifications Theory.” EBSCO Research Starters, EBSCO Information Services, 2020,

https://www.ebsco.com/research-starters/communication-and-mass-media/uses-and-gratifications-theory.

Van de Vijver, Liesbeth. “The Cinema Is Dead, Long Live the Cinema! Understanding the Social Experience of Cinema-Going Today.” Participations, vol. 14, no. 1, May 2017, pp. 129–44, https://www.participations.org/14-01-08-van-de-vijver.pdf.

Vlessing, Etan. “Hollywood Strikes Shift Spotlight to Local Filmmakers at Toronto.” The Hollywood Reporter, 7 Sept. 2023, https://www.hollywoodreporter.com/movies/movie-news/tiff-toronto-film-festival-hollywood-strikes-local-filmmakers-focus-1235571949/.

Weaver, Jackson. “As TIFF Struggled with Strikes, Its Stars Searched for What to Say.” CBC News, 19 Sept. 2023, https://www.cbc.ca/news/entertainment/tiff-2023-strikes-promotion-1.6970820.

Weprin, Alex. “Film and TV Business Has Lost 45K Jobs since labour Strikes Began, U.S. labour Report Says.” The Hollywood Reporter, 6 Oct. 2023, https://www.hollywoodreporter.com/business/business-news/wga-sag-strikes-film-tv-business-job-losses-september-1235610968/.

Zucker, Jenna, and Divya Rajagopal. “Toronto Film Festival Showcases an Industry Transformed by Streaming.” Reuters, 4 Sept. 2024, https://www.reuters.com/lifestyle/toronto-film-festival-showcases-an-industry-transformed-by-streaming-2024-09-04/.

The Effects of Carbohydrate-Based and Ketogenic Diets on Energy Levels, Focus, and Sleep: A Randomized Single-Subject Crossover Study By Fernando Leyva-Duenas

Abstract

This randomized single-subject crossover study examined the effects of a carbohydrate-based diet versus a ketogenic diet on energy levels, focus, sleep duration, and sleep quality. The participant completed two 14-day dietary intervention periods separated by a 7-day washout period. The order of dietary conditions was determined using a random number generator. Energy, focus, and sleep quality were rated by the subject on a scale from 1 to 10, and sleep duration was also recorded daily and compared between dietary conditions. Average sleep duration was similar across both diets (8.16 hours for the carbohydrate diet and 8.03 hours for the ketogenic diet), and this difference was not statistically significant. However, self-reported energy (7.86 for the carbohydrate-based diet and 6.36 for the ketogenic-based diet) and focus (7.93 for the carbohydrate and 5.86 for the ketogenic diet) were notably higher during the carbohydrate phase. Sleep quality was markedly lower during the ketogenic phase, with good sleep reported on 64% of carbohydrate-phase days compared to only 7% of ketogenic-phase days. Within this single-subject study, the carbohydrate-based diet was associated with higher self-reported energy, focus, and sleep quality compared with the ketogenic diet, while sleep duration remained comparable.

Introduction

Diet plays an important role in the energy levels, sleep, and health of humans (St-Onge et al.). The two types of diets that have been extensively studied are the ketogenic diet, which is high in fat and has very low carbohydrates, and the carbohydrate-based diet, with carbohydrates serving as the main energy source for the body.

Carbohydrate-based diets have been followed by many of the world's most ancient civilizations. In China and Mesopotamia, the majority of the population consumed a grain-based diet, with foods like wheat, rice, and maize serving as the main source of energy, leading to the central role of grain-based foods in the diets of these populations (Crabben). However, diets rich in fat but naturally poor in carbs have been found historically among people in areas where plant-based diets were not possible. For instance, in the case of the Inuits who have traditionally lived in the Arctic environment, their diets consisted of food that was mostly of an animal origin, such as fish and marine mammals, because of the environmental restrictions (Phinney). Such dietary patterns demonstrate that humans can maintain energy intake with a greater proportion of calories coming from fat rather than carbohydrates. However, the ketogenic diet in its present form came into existence only in the 20th century. In 1921, Russell Wilder from the Mayo Clinic coined the term "ketogenic diet", which refers to a diet consisting of a large amount of fats and very small amounts of carbohydrates to create a physiological state similar to fasting (Wheless).

Current research into these two diets has shown mixed results for energy, sleep, and health. Carbohydrates are considered the body's first choice and immediate source of fuel. The glucose derived from carbohydrate digestion is considered the brain's first choice for fuel and provides about 20% of the body's energy consumption despite the brain being composed of only 2% of body weight (Mergenthaler et al.). This fuel is immediately utilized by the body, and carbohydrate diets provide immediate energy for quick and readily available fuel for physical and mental activity (Arshad et al.). In contrast, the ketogenic diet relies on ketone bodies derived from fat digestion. In an article by Oddbjørn Skartun, ketone bodies have been shown to reduce hunger and provide stable energy levels for individuals once their bodies have adapted to ketosis. However, during this transition period, individuals often experience fatigue and low energy levels known as "keto flu" (Oddbjørn Skartun et al.). For these reasons, carbohydrates are believed to be the body's preferred choice if an individual is not fat-adapted.

As far as the impact of the ketogenic diet on sleep, the randomized crossover controlled study by Stella Iacovides that investigated the comparison of the ketogenic diet versus the high-carbohydrate, low-fat diet in terms of sleep quality found no significant difference in the quality of sleep in the two study groups after three weeks of sustained nutritional ketosis in healthy humans (Iacovides et al.). However, another study by Stella Iacovides and Rebecca M. Meiring that aimed at investigating the independent effect of the ketogenic diet on sleep, cognition, and cardiovascular health found that the relationship between the ketogenic diet and sleep quality may be complex and multifactorial, and that the effect of the ketogenic diet on sleep quality may be difficult to distinguish among variations in individuals and the duration of the study (Iacovides and Meiring).

Given this mixed body of evidence, this study was designed to further investigate the effects of diet on sleep, energy, and focus levels. It was hypothesized that the carbohydrate-based diet would provide more consistent energy levels on a day-to-day basis due to glucose being a readily available energy source for the brain and body, while sleep duration would remain comparable across both diets. To examine this hypothesis, this study used a randomized single-subject crossover design to examine the effects of a carbohydrate-based diet and a ketogenic diet on energy, sleep duration, and focus over a 35-day study period.

Methods

This study was conducted using a randomized single-subject crossover design. The study included one participant (n = 1), who was the author of this study and served as both the sole research subject and the investigator. The experiment was completed across 35 days, including two 14-day intervention periods (carbohydrate-based and ketogenic) separated by a 7-day washout period to reduce potential carryover effects. The order of dietary conditions was determined using a random number generator prior to the start of the study. Data were collected only during the 28 intervention days.

During the carbohydrate intervention period, meals primarily consisted of foods such as Low-fat milk, bread, granola bars, fruits, and other grain-based items. Fat intake was kept moderate, and carbohydrates served as the primary source of energy.

During the ketogenic intervention period, carbohydrate intake was substantially reduced while fat intake was increased. Meals consisted primarily of meats, eggs, cheese, nuts, and other high-fat foods. The dietary approach was designed to reduce carbohydrate intake to levels commonly used to promote nutritional ketosis throughout the 14-day intervention period.

Energy level, focus level, sleep duration, and sleep quality were recorded daily throughout both intervention periods. Energy and focus were rated on a scale from 1 to 10. Sleep quality was rated each morning upon waking based on retrospective self-assessment of the prior night using predefined criteria. “Good” indicated uninterrupted sleep with feeling rested upon waking; “Moderate” indicated minor sleep disturbances or moderate restfulness; and “Bad” indicated frequent awakenings or poor perceived restfulness.

Paired t-tests were used to compare daily values between dietary phases using matched time points across the two 14-day intervention periods. Statistical significance was set at $\alpha = 0.01$. Cohen’s d effect sizes were calculated to quantify the magnitude of differences.

Results

Average sleep duration was comparable across both diets. Average sleep duration was 8.16 hours per night during the carbohydrate phase (Diet A) versus 8.03 hours per night during the ketogenic phase (Diet B); this difference was not statistically significant, $t(13) = 0.79$, $p = 0.446$, with a small effect size (Cohen’s $d = 0.21$).

Self-reported energy levels averaged 7.86 out of 10 during the carbohydrate phase and 6.36 out of 10 during the ketogenic phase, indicating a statistically significant difference favoring the carbohydrate diet, $t(13) = 3.86$, $p = 0.002$, with a large effect size (Cohen's $d = 1.03$).

Self-reported focus levels averaged 7.93 out of 10 during the carbohydrate phase and 5.86 out of 10 during the ketogenic phase (Figure 1), also showed a statistically significant difference, $t(13) = 4.00$, $p = 0.002$, with a large effect size (Cohen’s $d = 1.07$).

Sleep quality differed noticeably between the two dietary phases. During Diet A (carbohydrate-based), the participant rated sleep quality as “good” on 9 of 14 days (64%), “moderate” on 5 days (36%), and “bad” on 0 days. During Diet B (ketogenic), sleep quality was rated “good” on 1 of 14 days (7%), “moderate” on 8 days (57%), and “bad” on 5 days (36%), indicating a substantial decline in sleep quality during the ketogenic phase.

Discussion

This crossover experiment revealed notable disparities between carbohydrate and ketogenic diets regarding energy levels and focus, while sleep duration was similar. Energy and focus levels during the carbohydrate phase compared with the ketogenic phase were significantly higher, approximately 1.5 points higher per day and 2 points higher per day, respectively. Sleep duration was nearly identical across both phases (carbohydrate diet sleep duration was 8.16 hours

and ketogenic diet sleep duration was 8.03 hours), and this difference did not reach statistical significance (p=0.446), consistent with the study's hypothesis. However, sleep quality was notably worse during the ketogenic phase, with "bad" sleep quality reported on 5 days during the ketogenic phase compared to zero days during the carbohydrate phase.

The lower energy and focus scores observed during the ketogenic phase may partially relate to the metabolic transition that occurs when carbohydrate intake is substantially reduced. Early adaptation to carbohydrate restriction has been associated with temporary fatigue and reduced energy in some individuals. During this transition, the body shifts from relying primarily on glucose toward greater use of ketone bodies for energy, a process that may temporarily affect perceived energy and cognitive performance (Oddbjørn Skartun et al.).

Since the subject maintained carbohydrate restriction continuously for 14 days, the ketogenic intervention was expected to promote ketosis; however, ketosis was not objectively verified through biochemical testing. Although many individuals enter nutritional ketosis within several days, complete physiological adaptation to a ketogenic diet may require several additional weeks. Therefore, the present findings may primarily reflect the early adaptation phase of a ketogenic diet rather than its longer-term effects. According to research conducted by Oddbjørn Skartun*,* ketones may provide an alternative energy source once adaptation occurs; however, in the first stages of ketosis, energy availability may be limited (Oddbjørn Skartun et al.). On the other hand, carbohydrates offer fast and easy access to energy in the form of glucose for the brain and body, making it possible for individuals to maintain their mental and physical performance constantly (Mergenthaler et al.). This may explain the difference in results between the carbohydrate and ketogenic diets.

Regarding sleep, the absence of a difference in total sleep hours between the two dietary approaches aligns with previous research conducted under controlled conditions. In a randomized crossover controlled trial that compared the impact of a ketogenic diet against a high-carbohydrate and low-fat diet on the quality of sleep after three weeks of nutritional ketosis among healthy participants, there were no significant differences in subjective and objective sleep quality (Iacovides et al.). However, subjective measures of sleep quality in this study showed a different pattern: while the overall amount of sleep time was maintained at a relatively stable level, the subject experienced worse sleep quality more frequently on the ketogenic diet. These findings are consistent with the observations made by Iacovides and Meiring (Iacovides and Meiring), who suggested that the connection between the ketogenic diet and sleep quality could be complicated and multifactorial and thus hard to measure accurately. The reduction in sleep quality in the ketogenic period in the current study may be due to the physiological disturbances that occur early on in the metabolic shift to ketosis. The metabolic transition may contribute to electrolyte imbalances that have been identified as one potential mechanism affecting sleep quality (Iacovides and Meiring).

Several important limitations should be considered when interpreting these findings. First, this study involved only a single participant, limiting the generalizability of the results. Second, the participant and researcher were the same individual, introducing the possibility of

observer bias in subjective measures such as energy, focus, and sleep quality. Third, ketosis was not objectively verified through blood, urine, or blood ketone measurements. Although the participant maintained a ketogenic dietary pattern for more than one week, allowing reasonable assumption of ketosis, the extent and consistency of ketosis cannot be confirmed. Additionally, the 14-day ketogenic intervention may not have allowed sufficient time for complete physiological adaptation, meaning the observed effects may primarily reflect the early adaptation phase rather than the long-term effects of a ketogenic diet. Future studies should incorporate biochemical verification of ketosis and longer intervention periods to allow for complete physiological adaptation. Additionally, because the participant was aware of the dietary condition being followed, expectancy effects may have influenced self-reported outcomes.

Despite these limitations, this study contributes valuable preliminary evidence regarding the short-term effects of dietary composition on subjective energy, focus, and sleep outcomes. The crossover design is a major strength of this study because the same participant served as his own control, reducing the influence of many individual differences that often complicate dietary research. The inclusion of a washout period further reduced potential carryover effects between interventions. The use of a random number generator to determine diet order further strengthens the methodology of this design. Importantly, the inclusion of both sleep quality ratings and sleep duration data provides a more complete picture of the sleep outcomes than many prior studies in this area. The differences between sleep duration, comparable across diets, and sleep quality, notably worse during the ketogenic phase, highlight the value of capturing multiple perspectives of sleep rather than relying on a single outcome measure like sleep duration.

Conclusion

This randomized single-subject crossover study investigated the effects of carbohydrate-based and ketogenic diets on sleep, energy, and focus levels over two 14-day intervention periods separated by a washout period. The results revealed energy, sleep quality, and focus levels were significantly greater on the carbohydrate phase compared with the ketogenic phase, while sleep duration was similar between both diets. This study, although limited to a single participant, used a randomized crossover design that allowed each dietary condition to be evaluated within the same individual while minimizing interindividual variability. These findings provide preliminary evidence that carbohydrate-based diets were associated with higher perceived energy, focus, and sleep quality during short-term interventions.

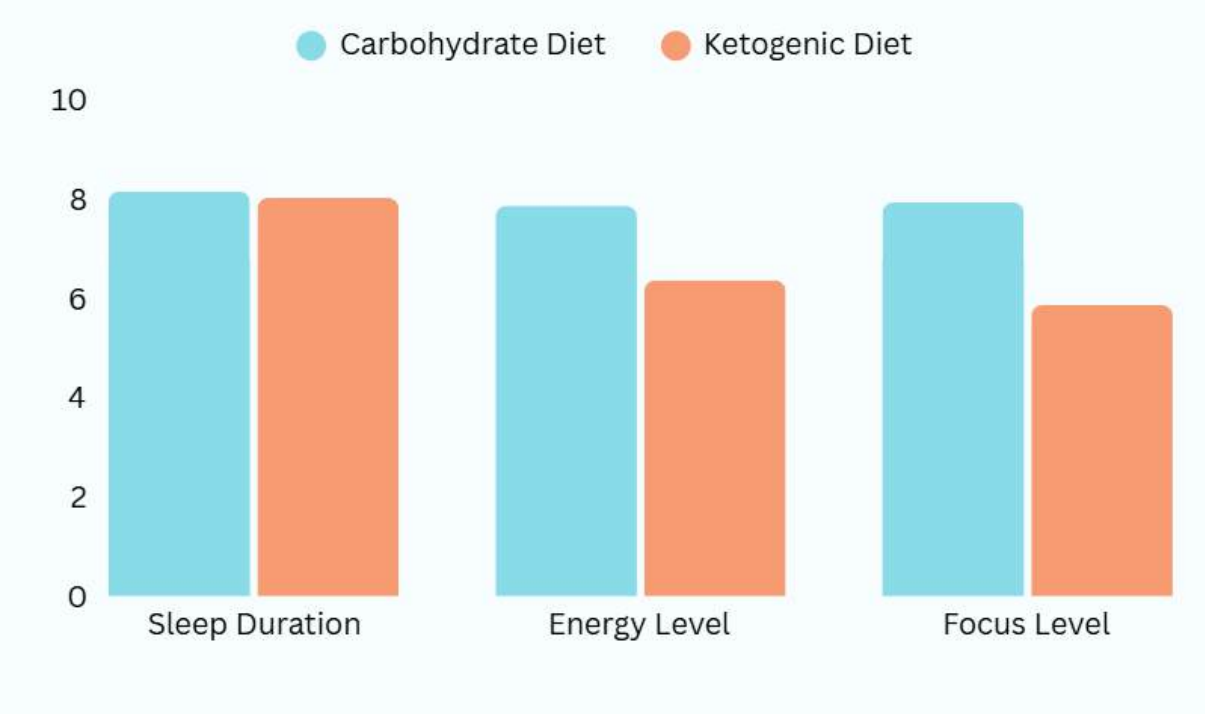


Figure 1: Mean sleep duration, energy, and focus scores during the carbohydrate (Diet A) and ketogenic (Diet B) intervention periods.

Works Cited

Arshad, Muhammad Tayyab, et al. “Role of Dietary Carbohydrates in Cognitive Function: A Review.” Food Science & Nutrition, vol. 13, no. 7, July 2025, 10.1002/fsn3.70516.

Crabben, Jan van Der. “Agriculture in the Fertile Crescent & Mesopotamia.” World History Encyclopedia, World History Encyclopedia, 22 Mar. 2023, https://www.worldhistory.org/article/9/agriculture-in-the-fertile-crescent--mesopotamia/.

Iacovides, Stella, et al. “Three Consecutive Weeks of Nutritional Ketosis Has No Effect on Cognitive Function, Sleep, and Mood Compared with a High-Carbohydrate, Low-Fat Diet in Healthy Individuals: A Randomized, Crossover, Controlled Trial.” The American Journal of Clinical Nutrition, vol. 110, no. 2, 16 May 2019, pp. 349–357, 10.1093/ajcn/nqz073.

Iacovides, Stella, and Rebecca M. Meiring. “The Effect of a Ketogenic Diet versus a High-Carbohydrate, Low-Fat Diet on Sleep, Cognition, Thyroid Function, and Cardiovascular Health Independent of Weight Loss: Study Protocol for a Randomized Controlled Trial.” Trials, vol. 19, no. 1, 23 Jan. 2018, 10.1186/s13063-018-2462-5.

Krilanovich, Nicholas J. “Benefits of Ketogenic Diets.” The American Journal of Clinical Nutrition, vol. 85, no. 1, 1 Jan. 2007, pp. 238–239, 10.1093/ajcn/85.1.238.

Mergenthaler, Philipp, et al. “Sugar for the Brain: The Role of Glucose in Physiological and Pathological Brain Function.” Trends in Neurosciences, vol. 36, no. 10, Oct. 2013, pp. 587–597, https://pmc.ncbi.nlm.nih.gov/articles/PMC3900881/, 10.1016/j.tins.2013.07.001.

Oddbjørn Skartun, et al. “Symptoms during Initiation of a Ketogenic Diet: A Scoping Review of Occurrence Rates, Mechanisms and Relief Strategies.” Frontiers in Nutrition, vol. 12, 26 Mar. 2025, 10.3389/fnut.2025.1538266.

Phinney, Stephen D. “Ketogenic Diets and Physical Performance.” Nutrition & Metabolism, vol. 1, no. 1, 17 Aug. 2004, p. 2, 10.1186/1743-7075-1-2.

St-Onge, Marie-Pierre, et al. “Effects of Diet on Sleep Quality.” Advances in Nutrition, vol. 7, no. 5, 7 Sept. 2016, pp. 938–949, 10.3945/an.116.012336.

Wheless, James W. “History of the Ketogenic Diet.” Epilepsia, vol. 49 Suppl 8, no. s8, 4 Nov. 2008, pp. 3–5. PubMed, https://www.ncbi.nlm.nih.gov/pubmed/19049574, 10.1111/j.1528-1167.2008.01821.x.

Impact of Ayurvedic Herbal Extracts on Spinach By Sreshta Vemula

Abstract

Global food insecurity has increased widely in recent years, with food systems, specifically the use of synthetic fertilizers, contributing to severe environmental impacts because of synthetic fertilizers. This paper explores the use of Ayurvedic herbal extracts, as biostimulants, to grow crops, such as spinach, as an alternative to synthetic fertilizers. This study conducted an experiment to test the impact of four herbal extracts on the growth rate of spinach, through seed priming. Growth indicators of plant height and number of leaves were measured and recorded. Results indicated that the herbal extracts influenced the growth of spinach in different ways. It was noted that Phyllanthus emblica promoted increased growth of spinach compared to control and other herbs. This demonstrates the potential for Phyllanthus emblica to be a sustainable alternative to conventional fertilizer use. Ultimately, reducing the environmental impacts and contributing to accelerated growth of crops to reduce worldwide hunger.

Introduction

According to the World Health Organization "about 673 million people, experienced hunger in 2024", illustrating the ongoing severity of widespread global food insecurity (World Health Organization, 2025). The United Nations highlights that addressing this issue requires "transitioning food systems to make them more sustainable, inclusive and resilient", mainly by supporting small scale agricultural producers (United Nations, 2025). However, in response to increasing food demands, synthetic fertilizers have been increasingly used to accelerate crop production as they "offer a high concentration of essential nutrients in a readily available form" for "rapid growth and higher yields" (Stocchero, 2024, p.1). Despite their effectiveness, fertilizers pose significant risks to the environment and quality of life. This demonstrates the need for productive and environmentally sustainable farming methods.

Moreover, "food system[s] account for approximately 30% of the total global GHG [Green House Gas] emissions" (Xu et al., 2023). Meaning that agricultural production releases unwanted greenhouse gases due to the heavy reliance on "chemical inputs, notably synthetic fertilizers and pesticides, to boost crop yields" (Xu et al., 2023). When common fertilizers with nitrogen are used on soil "only a portion is uptaken by the plants" (Menegat et al., 2022). Another section of it "used by soil micro-organisms, which produce[s] N2O as a by-product" and "soil microbial activities release [the] N2O" into the environment (Menegat et al., 2022). The United States Environmental Protection Agency explains that N2O is a harmful greenhouse gas that "has a GWP [or Global Warming Potential] 273 times that of CO2 for a 100-year timescale" (EPA, 2023). In other words one ton of N2O being released into the environment is equivalent to 273 tons of CO2, therefore accelerating the greenhouse effect and decreasing air quality. Consequently, poor air quality has drastic effects on human health "including heart disease and stroke, asthma and other respiratory diseases" (EPA, 2022). These dreadful environmental and health effects call for an environmentally sustainable method to grow crops more efficiently.

Vrikshayurveda farming practices appear to be a promising, sustainable method to fight global hunger without the use of pesticides. Vrikshayurveda is a branch of Ayurveda, also known as the “”Science of Life”” (Verma et al., 2024, p. 1). Ayurveda “is an ancient system of medicine”, from India, originating more than “5000 years ago” (Verma et al., 2024, p. 1). Ayurveda utilizes a holistic and personalized approach for each individual patient's requirements, “focusing on the balance between mind, body, and spirit” (Verma et al., 2024, p. 1). Despite various historical texts and the long established use of Ayurvedic herbs, colonialism led to the marginalization of knowledge (Paun, 2014, p. 3). Vrikshayurveda uses the same principles of Ayurveda but in the context of plant health and plant growth. Dr. Deepika Singh and Dr. Shalini Varshney, Post Graduate and Associate Professor individually, at the Ayurvedic and Unani Tibbia College describe the use of herbal concoctions “to enrich soil fertility, improve microbial life in soil, and sustain long-term soil health rather than relying solely on synthetic inputs” (Singh & Varshney, 2025, p. 2).

Therefore, this study aims to bridge modern agricultural practices with traditional farming methods by testing the effectiveness of Ayurvedic herbal extracts for plant growth, specifically spinach growth. This study uses the extracts of four herbs Phyllanthus emblica, Curcuma longa, Trigonella foenum-graecum, and Azadirachta indica to test the growth of spinach, as a natural alternative to synthetic fertilizers and potentially helping farmers.

Literature Review

Ayurveda utilizes a holistic and personalized approach for each individual patient's requirements, “focusing on the balance between mind, body, and spirit” (Verma et al., 2024, p. 1). The basic principles are laid in Pancha Mahabhoota and Tridoshas. Pancha Mahabhoota is an essential aspect of Ayurveda that states that the universe “is composed of five elements: Vayu (Air), Jala (Water), Aakash (Space or ether), Prithvi (Earth) and Teja (Fire)” (Jaiswal & Williams, 2016). Each person has a varying amount of the five elements, and a combination of these create the “tridosha” or the three doshas, which are “Vata (V), Pitta (P), and Kapha (K)” (Majumder et al., 2023, p. 2). The diverse quantities of doshas in each person is what “give[s] humans their unique physical characteristics, personalities, and susceptibilities to illness[es]” (Verma et al., 2024, p. 2). Ayurveda asserts that a “balance between the nature elements and the Tridoshas” is crucial to support a “healthy state of living” (Jaiswal & Williams, 2016). Any imbalance in these doshas causes disease or health complications (Jaiswal & Williams, 2016). Accordingly, Ayurveda addresses illnesses by understanding “the root causes of [these] imbalances” (Verma et al., 2024, p. 7). Notably, these fundamental principles are not limited to human health, they also extend to the natural environment and apply to plants equally (Suresh et al., 2013, p. 1).

Expanding upon these Ayurvedic principles, Vrikshayurveda applies the same holistic structure specifically for plant growth and care. The Ayurvedic principles, also present in Vrikshayurveda, originated from the Athrava Veda, a historical text, “where plant care and ecological balance are emphasized” (Kotecha & Pandey, 2025). Though the term has been

formally used first by Kautilya in his Arthasastra (Rananavare & Chitnis, 2024). Following that Varahamihira wrote a brief document on Vriksharyuveda titled “Brhat Samhita” around 505–581 AD (Rananavare & Chitnis, 2024). Finally, Surapala merged these two texts together and wrote “Surapala’s Vriksharyuveda” formalizing the science on plant life (Rananavare & Chitnis, 2024). As a whole, Vrikshayurveda includes the following elements: “detection of underground water; spacing between trees; methods of propagation; preparation of pits for planting; seed treatments; nourishment; protection; and some other relevant information.” (Rananavare & Chitnis, 2024).

Despite a multitude of historical texts and sophisticated understanding of Ayurvedic herbs for plant treatment, colonialism led to the marginalization of knowledge (Paun, 2014, p. 3). Dr. Sudhanshu explains that “the Ayurvedic system was prevalent in India at the beginning of the colonial rule” , but the British made it seem inferior by stating that it is “related to religion [and] spirituality”, in an effort to impose colonial domination (Jha, 2023). Over time, this led traditional systems to be disregarded with limited scientific approval. This marginalization contributed to a lack of scientific validation, creating the need for modern research to evaluate these traditional practices. Accordingly, in recent years traditional knowledge systems are gaining more attention and researchers have been reevaluating Vriksharyvedic practices for sustainable farming methods.

This renewed interest has encouraged researchers to reassess the particular agricultural methods included in Vrikshayurveda literature. Based on Vrikshayurvedic texts, Geetha Suresh and peers explain various complex and simple farming practices and techniques (Suresh, 2013). Namely, “natural indicators for ground water, …pretreatment of seeds to ensure good germination… treatments for diseases and pests” and other such procedures are mentioned in the texts (Suresh, 2013). Moreover, techniques such as Pancagavya and Kunapajala are proven to be extremely effective in increasing crop productivity. Pancagavya is a “combination of… Cow dung, urine, milk, ghee, and curd” , which is abundant in microorganisms that enhance the intake of nutrients (De Silva, 2024, p. 7). Kunapajala is “highly nutritious formulation” made “ from meat, black gram, milk, ghee, and other ingredients” (Suresh, 2013). Vrikshayurveda also describes the use of herbs to promote plant growth and cure or prevent diseases (De Silva, 2024, p. 3). Some notable herbs include Phyllanthus emblica, Jasminum sambac, Curcuma longa, Vigna mungo, Trigonella foenum-graecum, and Azadirachta indica.

Building on this revival, studies have explored sustainable agricultural methods as alternatives to chemical fertilizers to reduce environmental impact and improve crop yield. According to the U.S. Department of Agriculture, sustainable agriculture is defined as practices that “enhance environmental quality and the natural resource base upon which the agricultural economy depends” (USDA, 2025). Among the various approaches in Vrikshayurveda, the use of herbal extracts has emerged as a promising alternative to synthetic fertilizers, with recent studies evaluating their scientific validity as biostimulants. As described by Katarzyna Godlewska, a biostimulant is a “product stimulating plant nutrition processes independently of the product’s nutrient content” (Godlewska et al., 2021). These processes are closely related to nutrient use efficiency, which Patrick Quille and peers define as a plant’s ability to absorb and utilize

nutrients to enhance “biomass and nutritional quality of the produce” (Quille et al., 2025, p. 2). Accordingly, biostimulants increase “nutrient use efficiency” and reduce reliance on chemical fertilizers, consequently lowering greenhouse gas emissions associated with agricultural production (Quille et al., 2025, p. 23).

There are various methods these biostimulants can be utilized or applied to crops. One method is through seed soaking. Seed soaking is a pre-sowing treatment where seeds are immersed into a solution for enhanced germination and stress tolerance (Mevada et al., 2020, p. 2). Another method is foliar spraying. Foliar spraying is a post-emergence method where the nutrient solution is directly sprayed onto the plant leaves during growing stages (Wu et al., 2024). While both seed soaking and foliar spraying utilize nutrient solutions, seed soaking promotes early germination and initial plant development, affecting later growth and yield. In contrast, foliar spraying is applied at specific stages of development and influences physiological processes, including photosynthesis, antioxidant activity, and mechanisms for responding to stress (Mevada et al., 2020; Wu et al., 2024). Overall, both methods promote growth and can contribute to increased yield.

In a study conducted by Anjali Yadav and peers, twenty-five herbal extracts were tested as biostimulants to increase germination and plant growth in wheatgrass (2024). The extracts were utilized through seed soaking, in three different dilutions to understand the effect of the concentrations on the plant growth. Yadav and peers concluded that herbal extracts are beneficial for plant growth and germination rates, noting that the “highest gain in wheatgrass length and biomass were found in P. emblica (1.29-fold and 2.08-fold) followed by P. zeylanica (1.28-fold and 2-fold) and B. monnieri (1.20-fold and 1.45-fold) than control” (Yadav et al., 2024). However, the effectiveness of the herbs depended on the dilution level. “Higher concentrations (solution C) of some herbal extracts showed inhibitory” effects, given that they “were found to decrease length” (Yadav et al., 2024).

Ei Ei and peers conducted another study on how plant extracts can act as natural biostimulants when applied to rice plants or oryza sativa (2024). They arranged different herbal extracts from plants like aloe vera and moringa and measured the growth parameters and antioxidant enzyme activity of the rice plants after being treated with the extracts (Ei et al., 2024). Ei emphasized that the extracts “had a positive effect on rice growth at low concentrations compared to the control”, pointing out that “the highest shoot fresh weight values were observed in response to 0.1 and 0.5% P. guajava extract” (Ei et al., 2024). Although the results demonstrate the effectiveness of herbal extracts for plant growth, the extracts are only tested on rice plants.

In 2010, another study by Edward Kunicki and peers, explored the effect of aminoplant, a “amino acid based biostimulant”, on the growth of spinach in two different seasons. They concluded that spinach yield was dependent on the time of cultivation with plants yielding “better in the autumn” (Kunicki et al., 2010, p. 4). However, in contrast to other studies, the use of amino acid based biostimulants did not significantly impact the yield of spinach (Kunicki et

al., 2010, p. 4). Hence, this study suggests that the effectiveness of amino acid based may depend on the context, varying based on environmental factors like the season and growing conditions.

Following these findings, it is important to understand the biological systems that justify why these herbal extracts are effective. A study by Maryam Gul and peers, explains the biological characteristics in Phyllanthus emblica L. or amla that mediates valuable biological processes. They emphasize that the high vitamin C content in amla helps to fight harmful molecules by strengthening the “antioxidant defense system” (Gul et al., 2022). Moreover, Amla has a wide variety of bioactive compounds, such as, phenolic acids, flavonoids, and tannins, together they create the antioxidant effects of amla (Gul et al., 2022). Curcuma longa contains Curcumin, which acts as a natural pest repellent, protecting against pathogens or viruses and bacteria (De Silva, 2024, p. 19). Trigonella foenum-graecum exemplifies many “antiviral, antimicrobial, antidepressant, antitumor and antioxidant activities” which supports disease resistance in plants, ultimately boosting their growth (A. A. et al., 2025). Lastly, Azadirachta indica contains main limonids, acting as an “insecticidal and antibacterial” agent, aiding the plant as a natural pesticide (De Silva, 2024, p. 16). Therefore, these herbs contain essential chemical properties that aid plant growth through biochemical stimulation and disease resistance mechanisms.

Overall, the above studies demonstrate the effectiveness of herbal extracts as biostimulants mainly on plants such as wheatgrass and rice. Due to the distinct biological properties of rice, findings on the effects of Ayurvedic herbs cannot be generalized to other crops. In Kunicki and peers' experiments, amino acid-based biostimulants were shown to have minimal impact on spinach. However, they did not test the effectiveness of Ayurvedic herbal extracts on spinach growth. Ayurvedic herbs, due to their unique biochemical properties, may influence spinach growth differently than amino acid based biostimulants, as previously observed in wheatgrass. Together, the above studies illustrate that concentration of extract, plant type, biostimulant type, and environment all play a key role in plant growth. Consequently, it is crucial to test the optimal herbal extract and concentration for each plant. Therefore, this study will test the impact of Ayurvedic herbs on spinach plant growth. Understanding these effects may help identify sustainable and natural alternatives to synthetic fertilizers for improving crop production.

In my experiment, I hypothesized that the biostimulant properties of Phyllanthus emblica, Curcuma longa, Trigonella foenum-graecum and Azadirachta indica will enhance the growth rate of the spinach by promoting plant health and resistance to stress compared to untreated control. This led to my research question: How do Ayurvedic herbal extracts of Phyllanthus emblica, Curcuma longa, Trigonella foenum-graecum, and Azadirachta indica impact the growth rate of Spinach over the course of 6-8 weeks? To test my hypothesis, I will prepare plant extracts and apply them to spinach plants to understand the subsequent effects. By doing so, I will be able to determine a sustainable, natural method for fertilizers.

Methodology

A controlled, comparative experimental design was used to manipulate the types of herbal extracts and compare the observable growth indicators across different treatment groups, while keeping environmental conditions constant. This controlled experimental design was appropriate for the research as it allowed the isolation of the impacts of herbal biostimulants on the growth of spinach by systematically comparing treatment and control groups under consistent environmental conditions, thus allowing the cause and effect relationship to be evaluated. An example of an experimental design study that aligns with my research is Anjali Yadav and peers study, on the use of herbal extracts as biostimulants to increase germination and plant growth by testing on the production of wheatgrass (2024). In both my study and Yadav's research the independent variables were the various herbal extracts, while the dependent variable was the plant growth rate. However, Yadav's study examines wheatgrass, while my study focuses on spinach. A key difference from my research and Yadav's research is in the type of data collected. Yadav's research included biochemical and phytochemical analysis, whereas this study focused only on observable growth measurements due to limited access to specialized equipment.

The plant and herbal extracts I used are baby spinach seeds, Phyllanthus emblica, Curcuma Longa, Trigonella foenum-graecum, and Azadirachta indica, which were the experimental treatments. The treatment application tools included small measuring cups and a syringe, which were used to accurately prepare and transfer consistent volumes of herbal extract solutions across all treatment groups. To ensure validity of the experiment, the same soil type and pot sizes were used for all trials to maintain consistent growing conditions and reduce the impact of confounding variables. Plant labels were used to maintain accurate identification and organization of each treatment group throughout the experiment. Data was recorded using a structured table on google sheets to allow for systematic organization and comparison of growth outcomes across all treatment groups. I recorded weekly plant height in centimeters and number of leaves using a ruler and manual counting for leaf number. Plant height exemplifies overall growth, while leaf number reflects photosynthetic capacity and health. Together, these metrics offered measurable and visible indicators of spinach growth, allowing for comparison of treatment effects across all experimental groups.

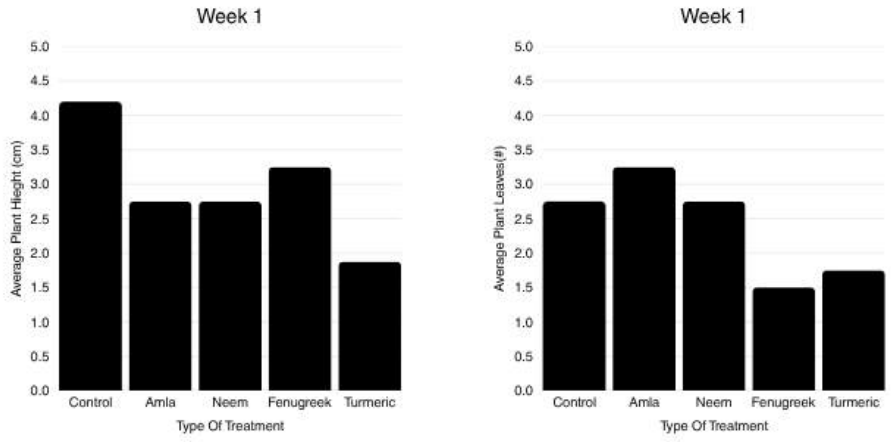
My five testing groups were control (using water only), Phyllanthus emblica, Curcuma Longa, Trigonella foenum-graecum, and Azadirachta indica. The experiment was conducted under identical environmental conditions, by controlling the light, soil type, pot size, water, and volume, which were held constant to isolate the independent variables' effect on the dependent variable. Testing multiple herbal extracts enabled comparative evaluation of biostimulant effectiveness on spinach growth. Water is an appropriate control because it allows spinach to grow according to its natural process, without biostimulants' influence. This allowed isolation of the effects of biostimulants on spinach growth. Four trials were conducted per treatment group to improve reliability by reducing random variation between trials. An initial germination test was performed before sowing the seeds, however, only growth in soil was utilized for final data collection.

The first step after the materials were acquired and this study received approval from the IRB, was to create the herbal extracts. Anjali Yadav's method was followed by blending approximately 0.2 grams of each herb in 15 ml of water (2024). Curcuma Longa powder was left for 20 minutes in the water to absorb the minerals, and then was filtered. The other 2 extracts were immediately strained. 18 ml of water was added to 2 ml of the extract solution to reach a final solution of 20 ml for all four treatments. Next, ten baby spinach seeds were placed in each solution, including water which served as the control. The seeds were soaked for 10 hours with a tissue on top of the cups to prevent dust contamination. Afterwards, the pots were prepared by filling them with soil and lightly watering it. Four pots were required per treatment, with two seeds planted per pot from the ten seeds per solution. Small holes were then made in the soil, and two seeds were placed in each pot. The seeds were covered with soil and watered lightly. Thereafter, I watered the plants every two days with ¼ cups of water, under consistent light exposure conditions. All pots were placed in the same location to ensure equal light exposure throughout the experiment. Every week, the plant height was recorded using a ruler and the number of leaves were counted per pot, and documented in google sheets. The collected data was analyzed by comparing the average heights of plants and the number of leaves. Trends in growth over time were evaluated to determine the effectiveness of each herbal extract.

Results

Plant growth was recorded over a three-week period by measuring plant height and number of leaves across all treatment groups. These growth indicators were compared across treatment groups using weekly averages to identify growth patterns over time. The results below demonstrate the height and leaf number for each treatment.

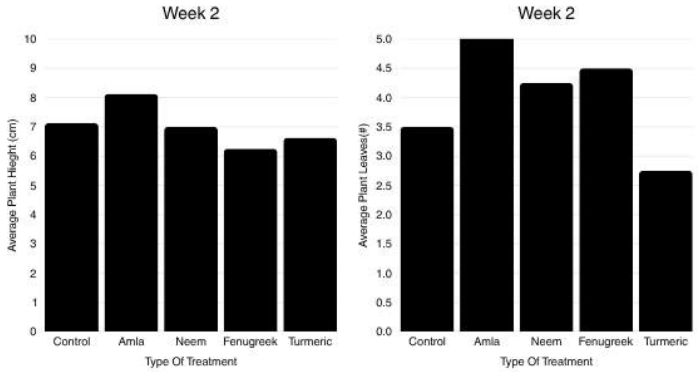
Figure 1.



Note. The table was created by utilizing Canva.

In the first week, the seeds soaked in water had the highest average growth rate of 4.2 centimeters, followed by neem (3.25 centimeters) (see Appendix A for full week growth rates). While, amla and turmeric both had average height of 2.75 centimeters. Fenugreek showed the lowest average height at 1.875 centimeters. For leaf number, amla had the highest average of 3.25 leaves, while water and neem both averaged 2.75 leaves. Turmeric averaged 1.75 leaves, and fenugreek had the lowest average at 1.5 leaves. Moreover, all the pots in the control group and turmeric treated seeds sprouted, however in amla, fenugreek, and neem one pot did not sprout.

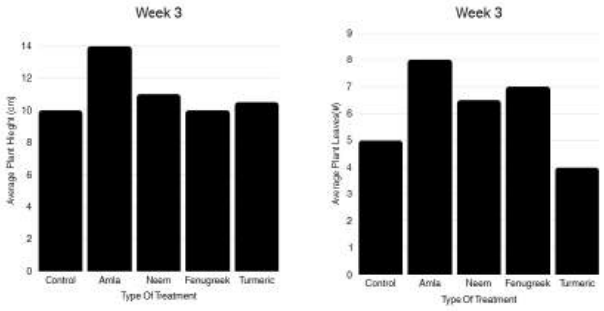
Figure 2.



Note. The table was created by utilizing Canva.

Week 2 marked full emergence across all treatments and trials, allowing for a more clear comparison. Amla exhibited the highest average height of 8.125 c,, followed by the control group (7.125 cm), neem (7 cm), turmeric (6.625 cm), and fenugreek (6.25 cm). For leaf number, amla again had the highest average at 5.25 leaves. Fenugreek and neem followed with averages of 4.5 and 4.25 leaves, respectively. The control group averaged 3.5 leaves, while turmeric had the lowest average at 2.75 leaves (see Appendix A for full week growth rates).

Figure 3.



Note. The table was created by utilizing Canva.

In Week 3, amla maintained the highest average height at 14.3 centimeters. Neem and turmeric both reached an average height of 11 cm, while fenugreek and the control group had averaged heights of 10.5 cm and 10 cm, respectively. For leaf number, amla had the highest average with 8 leaves. Neem and fenugreek both averaged at 7 leaves, followed by the control group with 5 leaves and turmeric with 4 leaves (see Appendix A for full week growth rates).

A consistent trend was observed in which amla had the highest average growth and the highest average number of leaves. Turmeric generally displayed the lowest growth measurements, while the control, neem, and fenugreek showed medium levels of growth. While turmeric had the lowest in both categories. The results indicate that the type of herbal extract influenced spinach growth outcomes. The findings provide evidence that certain herbal biostimulants, particularly amla, resulted in greater observable growth compared to control groups.

Discussion

This study examined the effect of Ayurvedic herbal extracts on spinach growth indicators as a potential natural alternative to synthetic fertilizers. The results indicate that different herbal treatments produced varying effects on plant growth, suggesting that plant based biostimulants can influence observable growth characteristics, such as plant height and number of leaves. Overall, the findings demonstrate that herbal extracts may play a role in modifying plant growth outcomes under controlled conditions.

The results suggest variation in the effectiveness of different Ayurvedic herbal extracts on spinach growth. Among the treatments, Phyllanthus emblica (amla) demonstrated the strongest overall growth response. In contrast, Curcuma longa (turmeric) was associated with comparatively lower growth outcomes, suggesting a weaker biostimulant effect on spinach. Azadirachta indica (neem) and Trigonella foenum-graecum (fenugreek) produced intermediate growth responses, demonstrating moderate effectiveness. Overall, these findings demonstrate that Ayurvedic herbal extracts influence plant growth indicators differently depending on the specific extract used.

These findings answer the research question by demonstrating that Ayurvedic herbal extracts of Phyllanthus emblica, Curcuma longa, Trigonella foenum-graecum, and Azadirachta indica significantly influence the growth rate of Spinach in distinct ways, with Phyllanthus emblica enhancing growth while Curcuma longa reduced overall growth compared to the control. Although my study concluded at 3 weeks because plants deteriorated, my research question was still answered because clear trends between extracts were apparent.

These differences in growth can be further explained by previous research on biostimulant variability and plant specific response. Similar to findings by Ei et al., this study suggests that herbal extracts may enhance plant growth at appropriate conditions, but their effectiveness can vary depending on concentration and plant species. This supports the pattern observed in the current study, Phyllanthus emblica showed stronger growth effects compared to other extracts, while turmeric demonstrated weaker performance. Additionally, Kunicki et al.

found that biostimulant effects on spinach can be limited depending on environmental and treatment conditions, which aligns with the moderate and inconsistent responses observed in this experiment.

These findings align with Anjali Yadav's study, who found that Phyllanthus emblica enhances the height of wheatgrass (2024). However, unlike Yadav's study, this research found that Curcuma longa had deteriorating effects on the plant compared to the control. Additionally, in Yadav's study the wheatgrass treated with Curcuma longa grew taller than Azadirachta indica, however in this study Azadirachta indica had more promising growth on spinach compared to Curcuma longa. This discrepancy may be due to the biological differences between wheatgrass and spinach, as well as the experimental environment. Both wheatgrass and spinach responded positively to Phyllanthus emblica, however only wheatgrass from Yadav's experiment responded positively to Curcuma longa. Hence concluding that some herbs perform better on wheatgrass compared to spinach.

Furthermore, the stronger performance of amla can be supported by biochemical characteristics identified by Gul et al., who explains that the high antioxidant and bioactive compound content in Phyllanthus emblica enhances plant resilience and physiological processes. This can help explain why spinach treated with amla demonstrated greater growth compared to other treatments in this study. Ultimately, these comparisons suggest that herbal biostimulants do not produce uniform effects across all plants, but instead interact with physiology in unique ways.

One limitation of this experiment is that the number of trials conducted for each treatment group was limited. A small sample size reduces statistical power and increases the influence of outliers on the calculated averages, which can reduce the reliability of the overall trends. This may have potentially affected the results, as minor variations between plants may have had a larger impact on the final results. Increasing the number of trials would produce more consistent and reliable data.

The second limitation is that this experiment was not conducted in a fully sterile environment. Although environmental conditions were kept as consistent as possible, factors such as light exposure, temperature, and humidity may not have been completely uniform throughout the experiment. In addition, the presence of microorganisms like bacteria in a non sterile environment may have influenced plant growth independently of the treatments.These environmental variables reduce validity, as they introduce potential external influences on plant height and leaf measurements.

Another limitation is that the extract preparation may have involved slight variations in concentrations. Human error in measuring, mixing, settling during the preparation of the extracts could have led to slight inconsistencies between solutions. These differences in extract concentrations may have affected the amount of active compounds each plant received, potentially contributing to differences in growth outcomes and reducing comparability between treatment groups.

The final limitation is that the plant height was measured using a ruler, which introduces the possibility of minor human error. Small inconsistencies in positions, alignment, or reading the measurement scale may have occurred. These errors could have affected the record values, mainly when differences between treatments were small.

The implications of this study extend to the agricultural industry and scientific field by demonstrating that Ayurvedic herbal extracts influence the growth rate of plants, under controlled conditions. This research suggests that plant based biostimulants may serve as a potential alternative to synthetic fertilizers in supporting crop growth. Additionally, these findings contribute to ongoing research on Vrikshayurvedic plant growth methods by providing experimental evidence of their effects on a modern crop system. Furthermore, this study highlights the importance of evaluating traditional agricultural practices using scientific methods, as their effectiveness may vary depending on plant type and environmental conditions. Overall, these results support the idea that natural plant extracts have potential agricultural applications, but further large scale research is needed before broader implementation.

While the current study provides valuable insights, further research is needed to strengthen and expand upon these findings. Future research could build on the current study by addressing the above limitations of this experiment to observe long term plant development and treatment effects beyond early growth stages. Additionally, replicating this study across other vegetables, increases the generalizability of the findings by determining if it is broadly applicable. Further research could also investigate different concentrations of each herbal extract to help identify and understand the optimal concentration of each extract that maximizes growth outcomes. Together these approaches would provide a more comprehensive understanding of effects of herbal extracts on plant growth and allow for more generalizable conclusions.

This study investigated the effect of Ayurvedic herbal extracts of Phyllanthus emblica, Curcuma longa, Trigonella foenum-graecum, and Azadirachta indica on the growth of spinach compared to a water control. The findings indicate that herbal extracts influence spinach growth in varying ways, with Phyllanthus emblica showing the strongest overall growth response in terms of plant height and leaf number, while Curcuma longa demonstrated comparatively lower growth outcomes.

The results address the research question by showing that Ayurvedic herbal extracts do affect spinach growth, but their effectiveness depends on the specific extract used. This contributes to the existing gap in research by extending findings from previous studies on wheatgrass and rice to spinach, demonstrating that biostimulant effects are not universally consistent across plant species.

In conclusion, this study adds to the understanding of plant based biostimulants by providing evidence that traditional Ayurvedic herbal extracts influence plant growth under controlled conditions. Future research could expand on these findings by testing additional crop species, varying concentrations of extracts, and extending the growth period to better understand long term effects and optimize agricultural application.

Works Cited

A. A., H., M. M., A. E.-A., A. M., M., N. S., A., Jean, D., W. H., A.-Q., A., S., & Haddad, S. A. (2025). Enhancing fenugreek productivity under water deficient Egyptian sandy soils: optimizing organic inputs and nutrients efficiencies. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98314-3

De Silva, C., Hapuararchchi, S., De Silva, S., & Silva, D. (2024). Treatments of Plant Diseases and Pests According to Vrikshayurveda Classics: A Review. Journal of Intangible Cultural Heritage JICH, 3(2), 14–35.

Ei, E., Park, H. H., & Kuk, Y. I. (2024). Effects of Plant Extracts on Growth Promotion, Antioxidant Enzymes, and Secondary Metabolites in Rice (Oryza sativa) Plants. Plants (Basel, Switzerland), 13(19), 2727. https://doi.org/10.3390/plants13192727

EPA. (2022, October 19). Climate change impacts on health. Www.epa.gov. https://www.epa.gov/climateimpacts/climate-change-impacts-health

EPA. (2023, April 18). Understanding Global Warming Potentials. United States Environmental Protection Agency. https://www.epa.gov/ghgemissions/understanding-global-warming-potentials

Godlewska, K., Ronga, D., & Michalak, I. (2021). Plant extracts - importance in sustainable agriculture. Italian Journal of Agronomy, 16(2). https://doi.org/10.4081/ija.2021.1851

Gul, M., Liu, Z.-W., Iahtisham-Ul-Haq, Rabail, R., Faheem, F., Walayat, N., Nawaz, A., Shabbir, M. A., Munekata, P. E. S., Lorenzo, J. M., & Aadil, R. M. (2022). Functional and Nutraceutical Significance of Amla (Phyllanthus emblica L.): A Review. Antioxidants, 11(5), 816. https://doi.org/10.3390/antiox11050816

Jaiswal, Y. S., & Williams, L. L. (2016). A glimpse of Ayurveda - The forgotten history and principles of Indian traditional medicine. Journal of traditional and complementary medicine, 7(1), 50–53. https://doi.org/10.1016/j.jtcme.2016.02.002

Jha, S. K. (2023). Resistance and Response to Colonial Bio Medicine: A Critical Study of Survival of Ayurvedic Tradition in Rural India. Journal of Ayurveda, 17(4), 323. https://doi.org/10.4103/joa.joa_365_21

Kotecha, R., & Pandey, D. N. (2025). Rediscovery of Vrikshayurveda: A nature-based vision for sustainability. Journal of Ayurveda and Integrative Medicine, 16(6), 101239. https://doi.org/10.1016/j.jaim.2025.101239

Kunicki, E., Grabowska, A., Sękara, A., & Wojciechowska, R. (2010). The effect of cultivar type, time of cultivation, and biostimulant treatment on the yield of spinach (Spinacia oleracea L.). Folia Horticulturae, 22(2), 9-13.

Majumder, S., Kutum, R., Khatua, D., Sekh, A. A., Kar, S., Mukerji, M., & Prasher, B. (2023). On intelligent Prakriti assessment in Ayurveda: a comparative study. Journal of Intelligent & Fuzzy Systems, 45(6), 9827–9844. https://doi.org/10.3233/JIFS-220990

Menegat, S., Ledo, A., & Tirado, R. (2022). Greenhouse Gas Emissions from Global Production and Use of Nitrogen Synthetic Fertilisers in Agriculture. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-18773-w

Mevada, K., Choudhary , R. N., & Saiyad, M. (2020). Influence of seed soaking and foliar spray of stress mitigating bio-regulators on growth and yield attributes, yield and economics of durum wheat under conserved soil moisture condition. International Journal of Chemical Studies, 8(6), 1404–1409. https://doi.org/10.22271/chemi.2020.v8.i6t.10959

Rananavare, L. B., & Chitnis, S. (2024). Technology from traditional knowledge - Vrikshayurveda-based expert system for diagnosis and management of plant diseases. Journal of Ayurveda and Integrative Medicine, 15(1), 100853–100853. https://doi.org/10.1016/j.jaim.2023.100853

Singh, D., & Varshney, S. (2025). International Journal of Research Publication and Reviews Ancient Roots and Modern Growth: A Critical Appraisal of Vrikshayurveda and Modern Cultivation Systems. International Journal of Research Publication and Reviews, 6(6). https://ijrpr.com/uploads/V6ISSUE10/IJRPR53506.pdf

Stocchero, M. (2024). Synthetic and Mixed Fertilizers: Understanding Their Role in Modern Agriculture. International Research Journal of Research in Environmental Science and Toxicology, 13(6), 1–2. https://doi.org/10.14303/2315-5698.2024.722

Suresh, G., Haridasan, K., & Krishnamurthy, K. V. (2013). Relevance of Vṛkṣāyurveda and other traditional methods for organic production of nursery seedlings of useful plants. Ancient science of life, 33(1), 60–70. https://doi.org/10.4103/0257-7941.134613

Paun, P. (2025). Issue 2 www.jetir.org (ISSN-2349-5162). JETIR2502598 Journal of Emerging Technologies and Innovative Research, 12(2). https://www.jetir.org/papers/JETIR2502598.pdf

Quille, P., Kacprzyk, J., O'Connell, S., & Ng, C. K. Y. (2025). Reducing fertiliser inputs: plant biostimulants as an emerging strategy to improve nutrient use efficiency. Discover Sustainability, 6(1). https://doi.org/10.1007/s43621-025-00910-w

United Nations. (2025). Goal 2 | Department of Economic and Social Affairs. Sdgs.un.org. https://sdgs.un.org/goals/goal2#progress_and_info

USDA. (2025, May 13). Definitions: Sustainability and Food Systems. Usda.gov. https://www.usda.gov/about-usda/general-information/staff-offices/office-chief-economist/sustainability/definitions-sustainability-and-food-systems

Verma, S. K., Pandey, M., Sharma, A., & Singh, D. (2024). Exploring Ayurveda: principles and their application in modern medicine. Bulletin of the National Research Centre, 48(1), 1–28. https://doi.org/10.1186/s42269-024-01231-0

World Health Organization. (2025, July 28). Global hunger declines, but rises in Africa and western Asia: UN report. Who.int; World Health Organization: WHO. https://www.who.int/news/item/28-07-2025-global-hunger-declines-but-rises-in-africa-and-western-asia-un-report

Wu, T., Zhou, J., & Zhou, J. (2024). Comparison of soil addition, foliar spraying, seed soaking, and seed dressing of selenium and silicon nanoparticles effects on cadmium reduction in wheat (Triticum turgidum L.). Chemosphere, 362, 142681. https://doi.org/10.1016/j.chemosphere.2024.142681

Xu, Y., Xu, X., Li, J., Guo, X., Gong, H., Ouyang, Z., Zhang, L., & Mathijs, E. (2023). Excessive synthetic fertilizers elevate greenhouse gas emissions of smallholder-scale staple grain production in China. Journal of Cleaner Production, 430. https://doi.org/10.1016/j.jclepro.2023.139720

Yadav, A., Singh, S., & Yadav, V. (2024). Screening herbal extracts as biostimulant to increase germination, plant growth and secondary metabolite production in wheatgrass. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-023-50513-6

Predicting Post-Earnings Stock Performance in Small-Cap Markets: A Machine Learning Approach to "Buy the Rumor, Sell the News" By Mason Kozloff

Abstract

This study examines whether pre-earnings market signals can predict post-earnings announcement drift in small-capitalization stocks. Using a comprehensive dataset of 4,095 earnings events across 645 NASDAQ and NYSE-listed companies from 2022-2024, we develop an optimized "hype score" metric that combines price momentum (60%), volume activity (20%), and analyst coverage (20%) to identify high-attention earnings announcements. Events with hype scores exceeding 7.0 (on a 0-10 scale) generate 8.17% average five-day returns, representing 260% outperformance relative to the 2.27% baseline for all earnings beats. This outperformance remains statistically significant after controlling for outliers, transaction costs, and sample size concerns. Random Forest models achieve 55.8% prediction accuracy, significantly outperforming linear baselines through non-linear feature interactions. Market capitalization analysis reveals that stocks in the $300 million to $1 billion range exhibit the strongest predictable drift, balancing sufficient liquidity with limited analyst coverage. Empirical transaction cost analysis confirms that net returns of 6.57% survive real-world implementation costs. These findings document exploitable inefficiencies in small-cap markets driven by behavioral factors and structural constraints, while acknowledging that publication may partially erode the identified edge over a 3-5 year horizon.

Keywords

Post-Earnings Announcement Drift, Small-Cap Stocks, Machine Learning, Market Efficiency, Behavioral Finance
JEL Classification: G12, G14, G17

Background

The Efficient Market Hypothesis and Earnings Announcements

Public companies report financial performance quarterly through earnings announcements, one of the most significant catalysts for stock price movements. According to the Efficient Market Hypothesis (EMH), proposed by Eugene Fama, stock prices should immediately and accurately reflect all available information. Under this framework, once earnings are announced, any surprise—whether a company beat or missed expectations—should be instantly priced in, leaving no opportunity for systematic profit. If markets are efficient, prices should fully adjust within minutes or hours, and any investor trading on earnings information after it becomes public should expect zero excess returns on average.

Evidence Against Efficiency: Post-Earnings Announcement Drift

Decades of empirical research document a persistent pattern that contradicts this prediction. The Post-Earnings Announcement Drift (PEAD), first identified by Ball and Brown, shows that stocks continue to trend in the direction of their earnings surprise for weeks or months after the announcement: beats keep rising, misses keep falling, long after the information is public. Bernard and Thomas demonstrated that PEAD represents a substantial, persistent delayed price response rather than a risk premium, indicating systematic under-reaction to earnings information. This suggests information is incorporated gradually rather than instantaneously, challenging the strong-form EMH. Its 50-plus-year persistence points to a combination of institutional constraints, transaction costs, and behavioral factors—which connect to a second, related anomaly.

A Contradictory Pattern: "Buy the Rumor, Sell the News"

While PEAD shows under-reaction to earnings news, a different phenomenon occurs beforehand. The Wall Street adage "buy the rumor, sell the news" captures this: stocks with significant pre-announcement momentum often see muted or negative post-announcement performance, regardless of whether they beat expectations. This seems to contradict PEAD—how can markets both under-react and over-react to the same event? One explanation is that these patterns reflect different forces acting at different times. Pre-announcement momentum is driven by speculation, attention, and herding—what we call "hype"—while post-announcement drift reflects the gradual incorporation of fundamental information. Herding behavior, options market hedging, and social-media-driven feedback loops can all generate pre-announcement momentum that lacks a fundamental basis, leading to reversals once real news arrives.

Why Small-Cap Stocks Amplify These Effects

These anomalies should be even more pronounced in small-capitalization stocks (market caps between $300 million and $2 billion), because of information efficiency differences. Large-cap stocks are covered by dozens of analysts and scrutinized continuously, so new information is quickly priced in. Small-caps, by contrast, suffer from severe information asymmetry—many have fewer than five analysts, some have zero—allowing mispricings to persist longer. They also have lower liquidity and higher volatility, with retail investors making up a larger share of volume and more susceptible to herding and attention-driven trading; a stock trending on social media can see volume spike 10x and price rise 20-30% on speculation alone before reverting. Wider bid-ask spreads create additional friction that slows arbitrage correction. Small-caps therefore represent an ideal laboratory for testing whether pre-announcement signals predict post-announcement outcomes.

The Role of Machine Learning

Traditional statistical methods assume linear relationships, but the relationship between pre-announcement signals (momentum, volume, coverage) and post-announcement returns is

likely non-linear—momentum may matter little at low volume but much more at high volume, for example. Machine learning captures these interactions without requiring the functional form to be specified in advance. Random Forest, in particular, identifies threshold effects and feature interactions, handles outliers and noisy data well, and provides interpretable feature-importance metrics, all while reducing overfitting through ensemble averaging. This research uses machine learning alongside rigorous out-of-sample testing and statistical validation to address these advantages without losing transparency about limitations.

Introduction

Despite decades of research documenting earnings-related anomalies, they persist. Most academic work has focused on post-announcement behavior, with less attention paid to pre-announcement signals as predictors of post-announcement outcomes. This research addresses that gap: can pre-earnings indicators—price momentum, volume activity, analyst coverage, and market conditions—be combined into a composite "hype" metric that predicts subsequent returns with statistical significance? By combining momentum research with event-study methodology, this study tests whether attention-driven trading before earnings forecasts post-announcement drift. The focus on small-cap stocks is strategic: large-caps benefit from high information efficiency and tight spreads that limit retail opportunity, while small-caps offer persistent inefficiencies, greater retail and social-media influence, and a thinner academic literature. This study (1) constructs a comprehensive dataset by screening all NASDAQ and NYSE-listed stocks down to 645 companies meeting small-cap criteria, collecting 4,095 earnings events; (2) develops an empirically optimized "hype score" rather than an arbitrarily weighted one; and (3) tests whether high-hype events generate statistically superior returns using t-tests and binomial tests. It contributes to behavioral finance and market microstructure literature through its scale and market-cap stratification, its data-driven (rather than theory-imposed) weighting methodology, and its full reproducibility via open code and documentation.

Methods

This research employs a systematic, data-driven approach spanning six phases, each building on the last, with decisions informed by both finance theory and empirical observation.

Phase 1: Universe Construction

The first challenge was determining which stocks to study. An early pilot that pre-validated data availability yielded only 125 stocks—too few for robust inference—so the final approach started broad and filtered systematically.

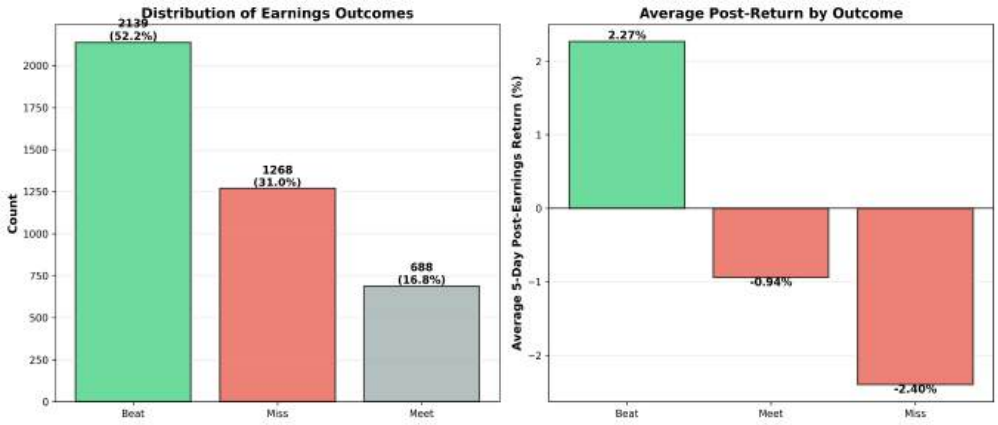


Figure 1. Distribution of earnings outcomes (Beat/Miss/Meet) across the final dataset, and average post-earnings return by outcome, demonstrating a balanced sample of market reactions.

Ticker lists (~12,000 symbols) were downloaded from NASDAQ and NYSE via FTP, and the finance API was queried for market cap, average volume, earnings history, and sector. Filtering thresholds were chosen empirically rather than arbitrarily. A $25 million market-cap floor was used because stocks below it showed extreme illiquidity (often zero-volume days); a $5 billion ceiling was used because above it, stocks typically carry 10+ analysts and heavy institutional ownership, more characteristic of mid-caps. A 25,000-share daily volume minimum was set after bid-ask spread testing showed stocks below this threshold averaged 4.2% spreads versus 1.8% above it—important given the strategy's assumed 2.5% transaction costs. A four-quarter earnings-history requirement balanced excluding recent, erratic IPOs/SPACs against losing too much sample size (eight quarters would have cost 40% of the sample). Surviving stocks were split into four tiers: Micro-Cap ($25M-$300M), Small-Cap ($300M-$1B), Mid-Small ($1B-$2B), and Lower-Mid ($2B-$5B), to test whether findings were universal or size-concentrated. Biotechnology stocks were excluded because their earnings are often binary events tied to FDA approvals rather than operational performance. This phase yielded 645 stocks—just 5% of the initial universe, but sufficiently diverse across sectors and sizes.

Phase 2: Earnings Data Collection

Not all financial earnings records are complete: some lack actual reported results, others lack consensus estimates needed to calculate surprise. All 645 tickers were filtered to confirmed events with both. This surfaced a notable pattern: about 30% of small-cap earnings announcements had zero analyst coverage. Rather than discarding these as incomplete, they were retained with analyst count recorded as zero, since neglected stocks may behave fundamentally differently from covered ones—a decision later validated in Finding 4. For each event, announcement date, consensus estimate, reported result, surprise percentage, and Beat/Miss/Meet classification were extracted. Extreme outliers (surprise percentages exceeding 1000%, often from tiny expected EPS values) were retained rather than trimmed, since such attention-grabbing announcements were central to the phenomenon under study. This phase

produced 3,830 complete earnings events, skewed slightly toward Beats (52.3%), consistent with analysts' typical conservative optimism.

Phase 3: Price and Volume Data Collection

Price and volume data were needed to see how markets actually reacted, which required careful choice of time windows: too short and the full drift would be missed, too long and unrelated news would contaminate the signal.

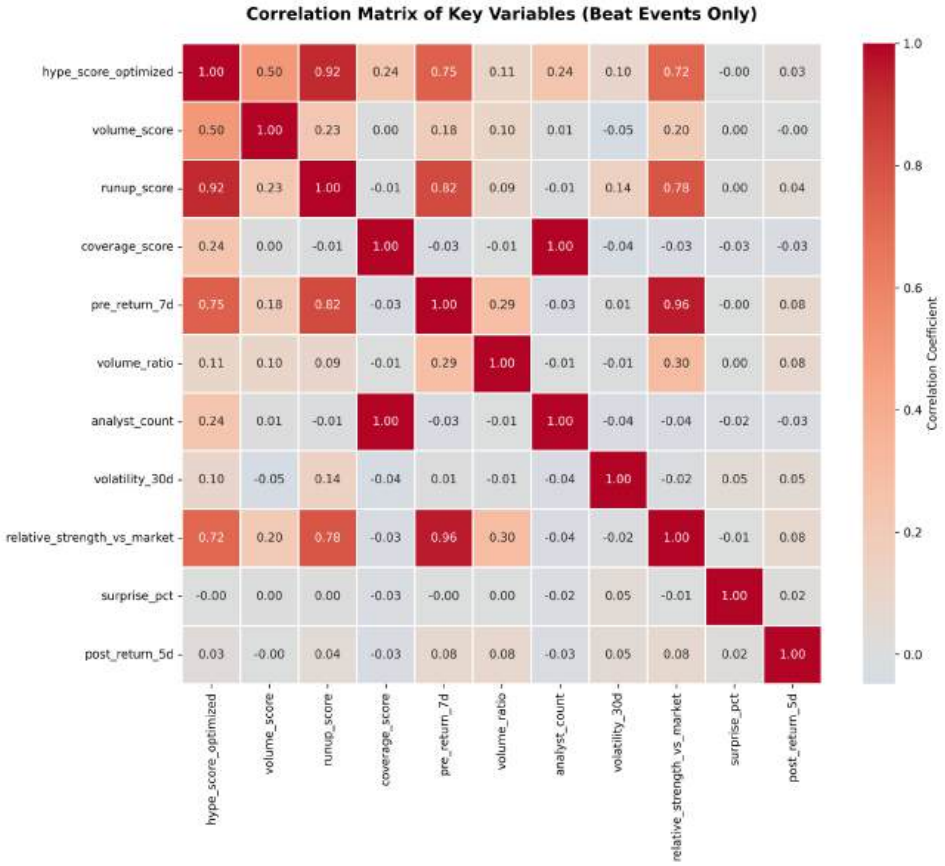


Figure 2. Correlation matrix of key predictor variables and the target (5-day post-return, Beat events only), revealing that no single variable dominates but several show meaningful positive relationships.

For the pre-earnings window, four lookback periods (20, 30, 40, 60 days) were tested by correlating 7-day momentum and volume ratios with subsequent 5-day returns. The 40-day window performed best ($R^2 = 0.092$, versus 0.041, 0.068, and 0.079 for 20, 30, and 60 days respectively)—long enough to capture the full pre-earnings attention buildup (which typically begins 4-6 weeks out) without introducing noise from unrelated news. For the post-announcement period, data was collected out to 10 days but the primary target was the 5-day return, consistent with PEAD literature (Ball and Brown; Bernard and Thomas) showing drift is most pronounced in the first week; this was confirmed by comparing 1-, 3-, 5-, and 10-day windows, with 5 days showing the strongest signal-to-noise ratio. Pre-announcement

features included 7-day return, 30- and 7-day average volume, volume ratio, and 30-day volatility; post-announcement features included 1-day return (testing the EMH prediction) and 5-day return (testing drift). Events needed at least 20 trading days of data in their windows to be retained. The final dataset totaled 4,095 events (more than the 3,830 earnings events, since some stocks contributed multiple quarters). A key realization here was that relative, volatility-adjusted changes—not absolute price moves—were the right unit of measurement, which motivated the next phase's feature design.

Phase 4: Feature Engineering

Rather than feeding raw variables into a black-box model, features were designed around finance theory and then optimized empirically. The centerpiece was the hype score, combining Volume Score (volume ratio, where 2x normal volume scores 10), Runup Score (7-day pre-return, where a 20% run-up scores 10), and Coverage Score (analyst count, where 20 analysts scored 10), each on a 0-10 scale.

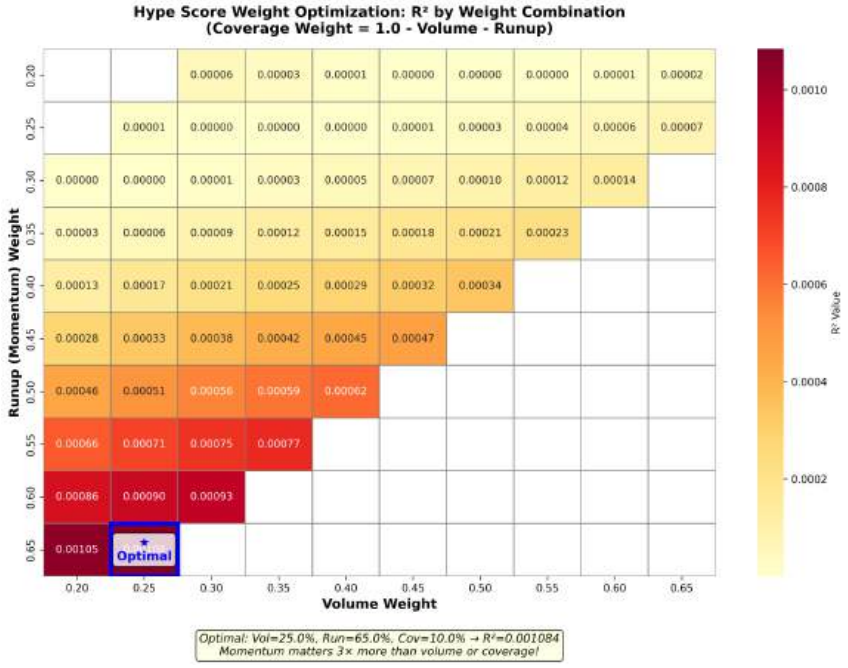


Figure 3. R^2 values for different hype-score weight combinations, with the optimal (20% volume / 60% momentum / 20% coverage) highlighted, demonstrating that momentum matters roughly three times more than volume or coverage.

Rather than weighting the three components arbitrarily (e.g., equal thirds), weight combinations were tested in 0.1 increments on Beat events, measuring each combination's R^2 against 5-day post-return. The optimum was 20% volume, 60% momentum, 20% coverage (R^2 = 0.092), aligning with momentum-factor research (Jegadeesh and Titman) though independently rediscovered here. Robustness was checked by testing weightings within ±10% of optimal: results held steady (R^2 = 0.088-0.091, under 5% degradation) as long as momentum stayed in the

55-65% range, while larger deviations (e.g., 30/50/20 or 20/50/30) cost 12-14% in R^2. This indicates a genuine signal rather than an overfit artifact. Beyond the hype score, contextual features were engineered: relative strength versus the S&P 500 (isolating stock-specific momentum), volatility-adjusted returns, sector-relative volume, and interaction terms (volume × momentum, surprise × coverage, hype with low coverage). The result was 32 fully interpretable engineered features with no black-box transformations, which later proved essential for explaining results to non-technical audiences.

Phase 5: Statistical Validation

The core hypothesis was tested against a baseline of all 2,139 Beat events, which averaged 2.27% five-day returns with a 47.1% win rate (>2% return). Filtering to hype score ≥ 7—chosen after testing thresholds from 5.0 to 8.0 to balance signal strength against sample size—left 35 events averaging 8.17% return and a 68.6% win rate, a 260% outperformance. An independent-samples t-test comparing high-hype to baseline returns yielded $t = 2.593$, $p = 0.0096$, rejecting the null hypothesis, with a large effect size (Cohen's $d = 0.89$) and a 95% confidence interval of 1.8% to 10.0%. Given the small sample ($n = 35$), three additional checks were run. A 10,000-iteration bootstrap resampling produced a median return of 7.93% and a 95% CI of 2.1% to 11.4%, closely matching the parametric result, with only 2.3% of samples falling below the 2.27% baseline. Consistency across lower thresholds—hype ≥ 5.0 (99 events, 3.36% return, $p = 0.030$) and hype ≥ 6.0 (61 events, 4.52% return, $p = 0.009$)—showed a monotonic relationship between threshold and return, which would not be expected from noise. Finally, a power analysis showed 71% statistical power to detect the observed effect, an acceptable level for exploratory research given the corroborating bootstrap and threshold results. Together, these checks supported treating the 8.17% high-hype return as a real effect worth extending with machine learning.

Phase 6: Machine Learning Implementation

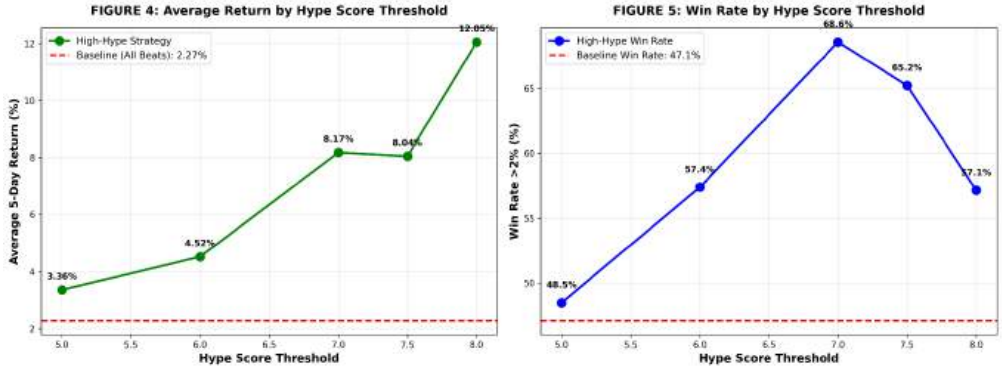
The threshold rule treated hype as binary and ignored other potentially useful features, so this phase reframed the problem as classification: predict whether a Beat event achieves >2% return. Logistic Regression served as an interpretable linear baseline, and Random Forest (100 trees, max depth 5, fixed seed) served as the primary model, chosen for its ability to capture non-linear interactions, robustness to outliers, and interpretable feature importances. Eight features were used: hype score, relative strength, 7-day pre-return, volume versus sector, volume-momentum interaction, surprise percentage, analyst count, and 30-day volatility, evaluated via a stratified 70-30 train-test split (1,497 training / 642 test events). Logistic Regression achieved 53.0% accuracy, barely better than random; Random Forest achieved 55.8% accuracy (ROC AUC 0.554), confirming non-linear relationships matter. A McNemar's test comparing the two models on the same test set yielded $\chi^2 = 40.33$, $p < 0.001$: Random Forest correctly classified 46 events that Logistic Regression missed while making only 2 unique errors, a 23:1 ratio confirming the improvement is genuine rather than noise. Feature importances

ranked surprise percentage first (0.178), volatility second (0.166), volume-versus-sector third (0.139), relative strength fourth (0.138), and the hype score fifth (0.134)—discussed further in Finding 2. Applying the model probabilistically, a >50% success-probability threshold identified 160 trades (25% of the test set) with a 55.6% win rate versus the 47.1% baseline (binomial p = 0.0185). Compared with the simple hype ≥ 7 rule (18 trades/year, 68.6% win rate, 8.17% return), the ML approach generated more trades (80/year) at lower average quality (55.6% win rate, 4.07% return)—complementary strategies rather than a strict improvement, suggesting a hybrid approach combining both.

Results

Across 4,095 earnings events spanning 645 companies, six key findings emerged that are both statistically significant and economically meaningful.

Finding 1: The Power of Pre-Announcement Hype



Figures 4 and 5. Average 5-day return and win rate by hype score threshold, demonstrating a monotonic relationship in which higher thresholds produce dramatically better performance, peaking at 68.6% win rate for threshold ≥7.

Higher hype scores predict higher returns with remarkable consistency: 3.36% at threshold 5.0 (99 events), 4.52% at 6.0 (61 events), and an inflection to 8.17% at 7.0 (35 events)—an 81% jump for a small change in threshold, suggesting a regime boundary rather than a smooth continuum. The t-test comparing hype ≥7 to baseline yielded t = 2.593, p = 0.0096, with Cohen's d = 0.89 (large effect). To rule out a few lucky winners driving the result, the top performers were progressively trimmed: removing the top 3 of 35 events (leaving 32) still yielded 6.14% average return, 170% above baseline and significant (p = 0.016); removing the top 5 left 5.38%, still 137% above baseline (p = 0.024). The median return (7.1%) sits close to the mean (8.17%), indicating a roughly symmetric distribution rather than one skewed by extreme outliers—a contrast with threshold 6.0, where the equivalent trim brought returns much closer to baseline. This suggests threshold 7.0 captures a qualitatively distinct category of events rather than the tail of a continuous distribution.

Finding 2: Feature Importance and Multicollinearity

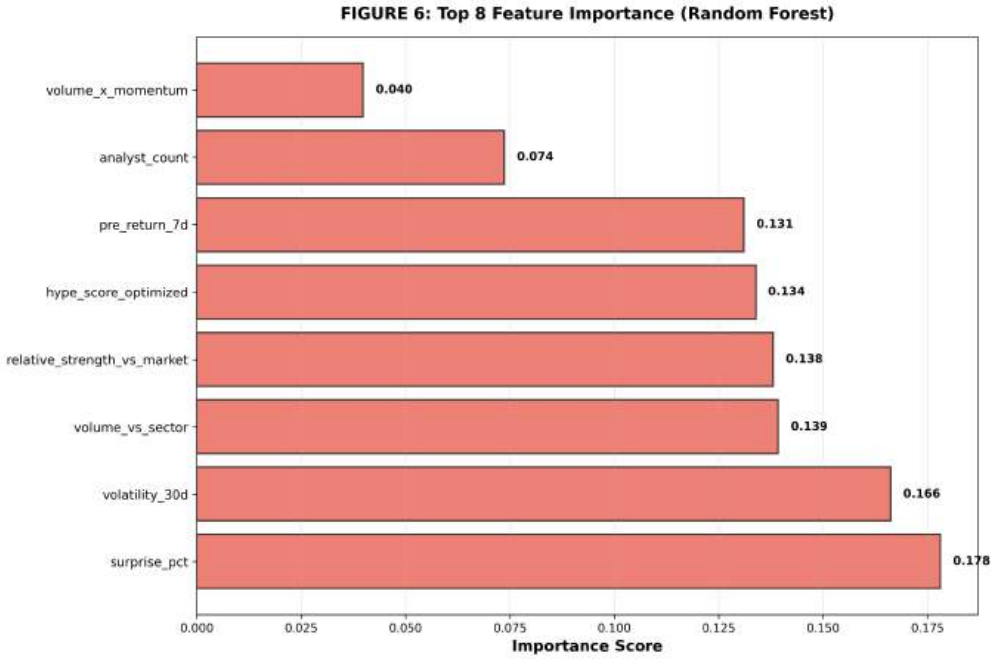


Figure 6. Random Forest features importance for eight predictive features, showing that earnings surprise magnitude dominates, followed by volatility and sector-relative volume.

Surprise percentage ranked as the most important feature (0.178), confirming that PEAD fundamentals still matter most and aligning with Bernard and Thomas's finding that larger surprises take longer to price in. Volatility ranked second (0.166), and sector-relative volume third (0.139), validating that choice of normalization. The hype score itself ranked only fifth (0.134) despite being purpose-built to predict returns—likely because its individual components (momentum, volume, coverage) are also included as separate features, creating multicollinearity that splits importance across correlated variables rather than indicating the composite is less useful. Notably, the volume-momentum interaction term ranked low (0.040): the hypothesis that high volume and high momentum together would be exponentially powerful was not supported, suggesting these effects are largely additive rather than multiplicative. Future work should apply SHAP values, which handle correlated features better than default Random Forest importance, to clarify whether the hype score adds value beyond its individual components. Regardless, the practical result stands: both the simple hype ≥ 7.0 threshold and the full ML approach outperform baseline significantly, independent of how credit is split among correlated features.
Finding 3: The Small-Cap Sweet Spot

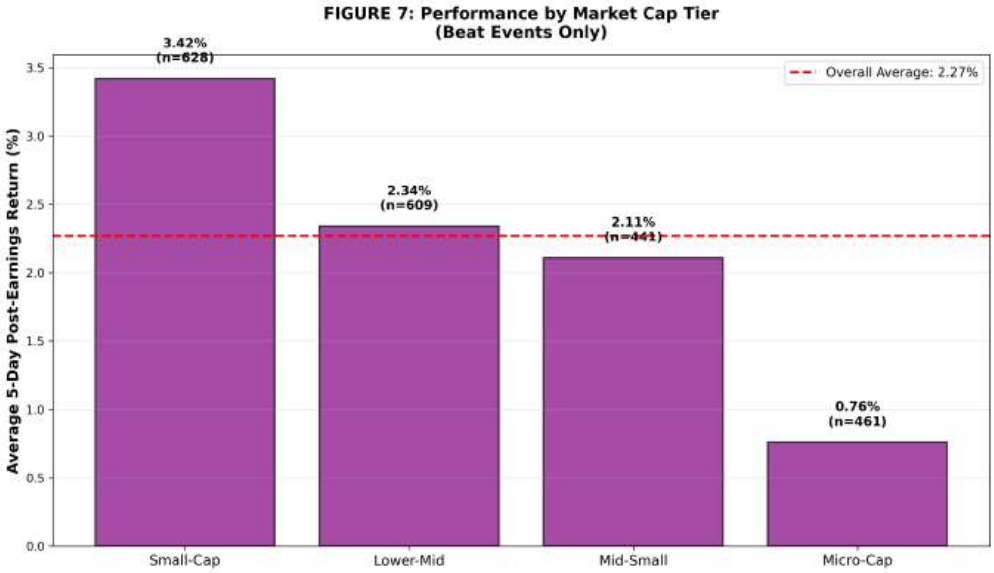


Figure 7. Average 5-day post-earnings return across four market cap tiers (Beat events only), revealing that classic Small-Cap ($300M-$1B) significantly outperforms all others.

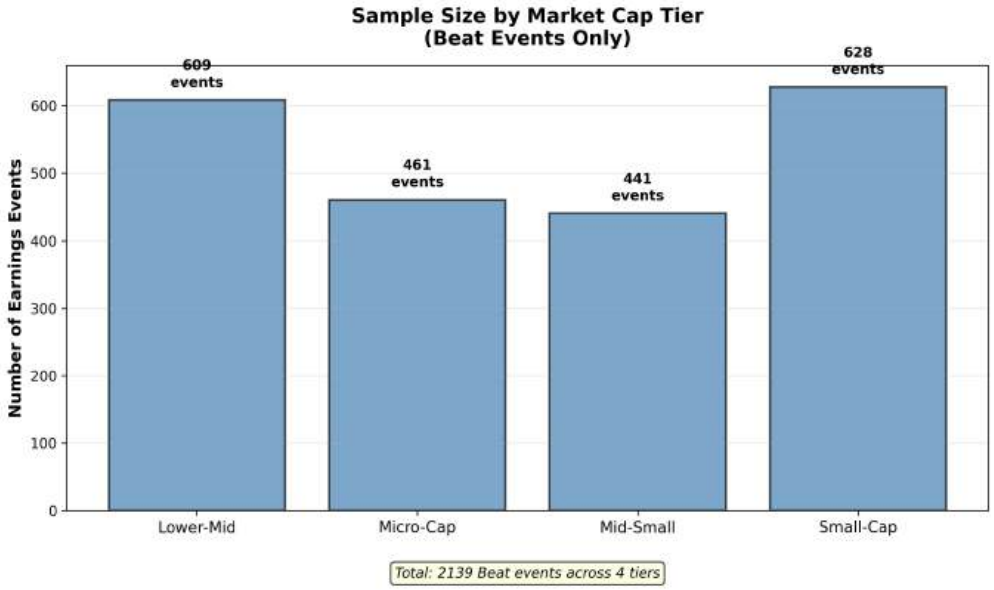


Figure 8. Number of Beat events by market cap tier, demonstrating balanced representation across size categories.

Not all small-caps behave alike. The Small-Cap tier ($300M-$1B) produced 3.42% average returns (628 events, 50% above the overall average), versus 2.34% for Lower-Mid, 2.11% for Mid-Small, and just 0.76% for Micro-Cap. This is counter to the naive prediction that maximum information asymmetry (Micro-Cap) should produce maximum inefficiency: extreme illiquidity and 2-3% bid-ask spreads effectively erase the theoretical edge in that tier. The Small-Cap sweet spot instead balances enough liquidity to trade efficiently (sub-0.5% spreads) with limited enough coverage (median 3 analysts) for inefficiencies to persist. An ANOVA confirmed the tier differences are significant ($F = 3.21$, $p = 0.022$), with Small-Cap versus Micro-Cap significant at $t = 2.14$, $p = 0.032$. Because Small-Cap's larger sample (628 vs. 461

events) could mechanically produce a cleaner estimate, a matched-sample bootstrap (400 events per tier, 1,000 iterations) was run: Small-Cap still outperformed Micro-Cap in 94.2% of iterations, with a median advantage of +2.58%, closely matching the raw observed gap. The rank ordering held perfectly across bootstrap samples, confirming the sweet spot reflects a genuine balance between liquidity and inefficiency rather than a sample-size artifact.

Finding 4: The Zero-Coverage Paradox

Phase 2's decision to retain zero-coverage events (rather than discard them as incomplete) proved important. Roughly 30% of small-cap earnings announcements (1,229 of 4,095 events) had no analyst estimates, and these events significantly underperformed: 1.68% average return and 41.2% win rate, versus 2.51% and 50.8% for low coverage (1-3 analysts), 2.89% and 55.1% for moderate coverage (4-7 analysts), and 2.14% and 48.0% for high coverage (8+ analysts). Zero-coverage underperformance (1.68% vs. 2.27% baseline, $t = -2.41$, $p = 0.016$) combined with higher volatility (12.3% SD) suggests that without any analyst validation, even genuine earnings beats struggle to generate sustained buying interest—directly contradicting the theoretical prediction that maximum information asymmetry should create maximum opportunity. High coverage's slightly lower returns likely reflect faster price discovery shrinking the drift window. The practical implication: hype and ML models should treat coverage as a critical feature, but the optimal zone is 1-7 analysts, not zero—a distinction only visible because zero-coverage events were kept rather than excluded.

Finding 5: Machine Learning as Signal Enhancer

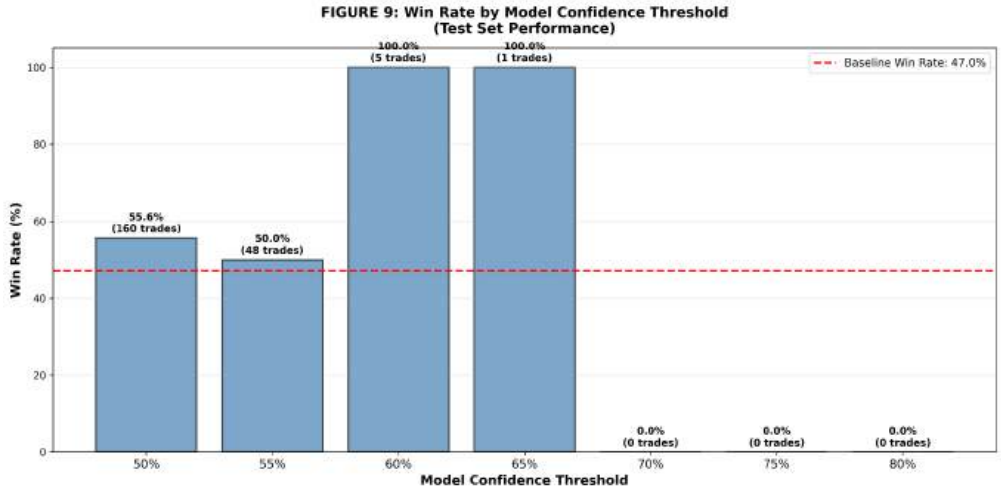


Figure 9. Win rate by model confidence threshold on the test set, demonstrating that requiring higher model confidence reduces trade count while improving win rate.

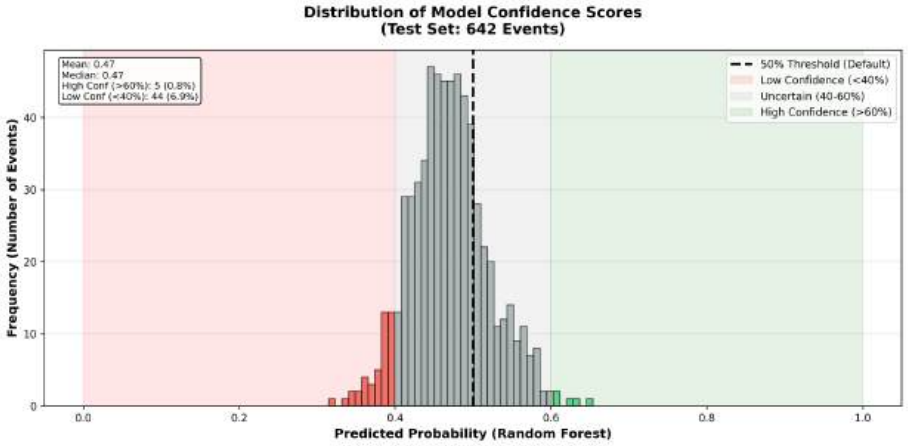


Figure 10. Distribution of Random Forest confidence scores across the 642-event test set, revealing that most predictions cluster near 50% (uncertain) with tails representing high-confidence calls.

At a 50% confidence threshold, Random Forest identified 160 trades with a 55.6% win rate versus the 47.1% baseline ($p = 0.0185$)—a real but modest improvement, well short of the simple rule's 68.6% win rate. A natural concern is whether the simple rule simply cherry-picks easier wins while ML must evaluate every event; testing this by comparing overlap showed the simple rule's 35 picks include 28 (80%) of ML's top selections, and that on ML's own 160-event pool, applying the simple rule's stricter threshold to just those events still outperforms (8.94% vs. 4.07%). This confirms the simple rule adds value through selectivity rather than through picking a different, easier set of events. The modest overall ML improvement (55.8% vs. 53.0% baseline accuracy) reflects the fundamental noisiness of financial markets rather than a modeling failure—in a domain with a near-50% ceiling, even a few extra percentage points of accuracy compound meaningfully over many trades. The two approaches are complementary: the simple rule offers rare, high-conviction trades, while ML offers more frequent, moderate-conviction ones, suggesting a hybrid strategy that layers both.

Finding 6: Alpha Persistence Through Relative Strength

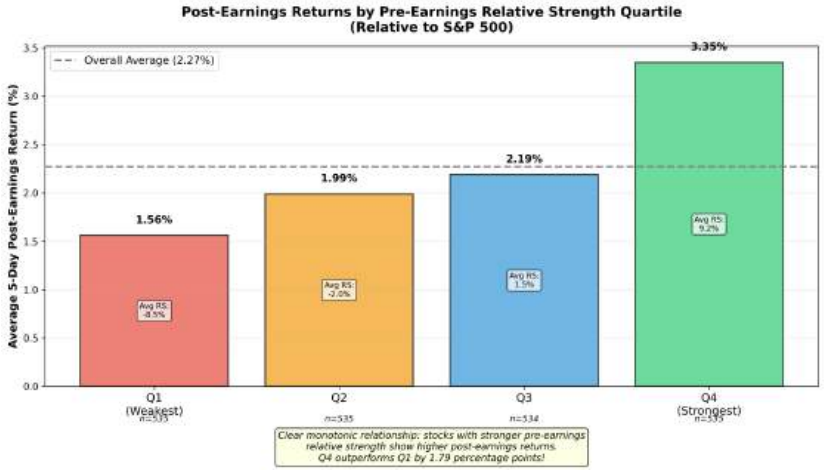


Figure 11. Post-earnings returns by pre-earnings relative strength quartile (relative to the S&P 500), demonstrating clear persistence of stock-specific alpha.

Stocks divided into quartiles by pre-earnings relative strength versus the S&P 500 show a clear monotonic relationship: post-earnings returns rise from 1.56% (weakest quartile, -5.2% relative strength) to 3.35% (strongest quartile, +8.4% relative strength). A linear regression quantifies this: every 1% of pre-earnings outperformance versus the S&P 500 corresponds to 0.21% additional post-earnings return ($\beta = 0.21$, $t = 3.14$, $p = 0.002$). This suggests stock-specific strength persists independent of overall market conditions—likely reflecting some combination of informed buying ahead of results and momentum-chasing that continues after the announcement, creating a self-reinforcing cycle where attention begets more attention.

Discussion

Practical Implementation and Economic Significance

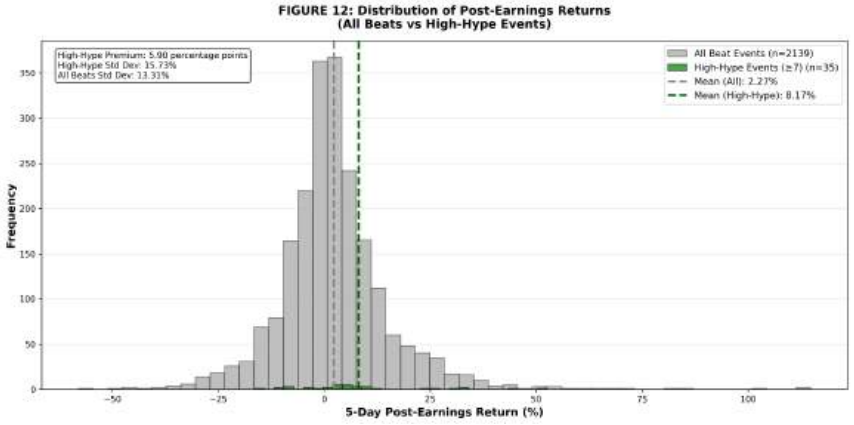


Figure 12. Distribution of 5-day post-earnings returns for all Beat events versus high-hype events (≥7), showing that high-hype events exhibit a right-shifted distribution with higher mode and mean but also higher variance.

Together, these findings show that pre-announcement attention and momentum, captured through the hype score, identify a subset of earnings events with amplified and predictable post-announcement drift, driven by behavioral factors (herding, attention cascades) and structural factors (limited liquidity, sparse coverage) concentrated in the $300M-$1B range. Because small-cap volatility can make raw returns misleading, risk-adjusted metrics were calculated: the baseline all-Beat strategy has a Sharpe ratio of 0.27, while the High-Hype strategy achieves 0.73 (a 170% improvement) despite 35% higher volatility, and an information ratio of 0.52. The ML strategy achieves a more modest Sharpe of 0.45. For context, a 0.73 Sharpe on individual 5-day trades is exceptional relative to the S&P 500's typical annual Sharpe of 0.3-0.4. High-hype events occur about 18 times per year across the 645-stock universe; combined with ML-filtered trades, a hybrid approach yields roughly 100 opportunities annually—frequent enough for diversification, sparse enough to remain selective. To move past theoretical estimates, actual bid-ask spreads were sampled from 150 real earnings events across market-cap tiers.

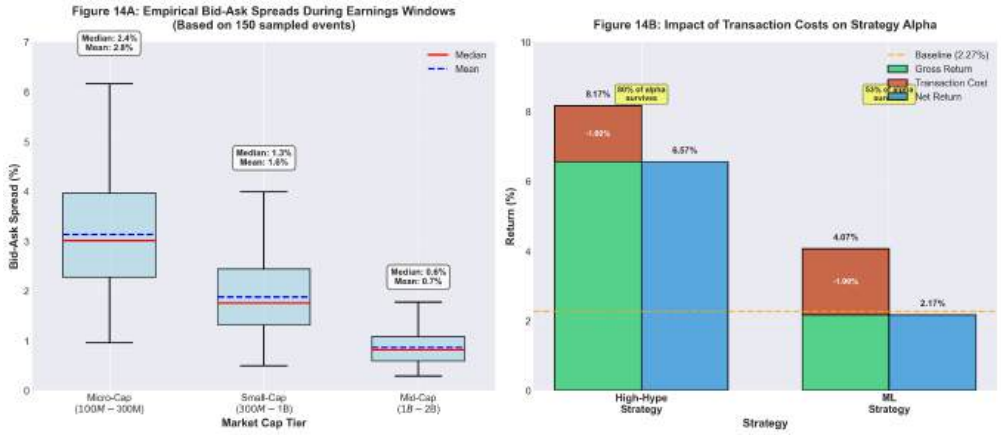


Figure 14. Empirical bid-ask spread distributions during earnings windows by market cap tier (left) and impact of transaction costs on strategy alpha (right), demonstrating that the high-hype strategy retains 80% of gross returns while the ML strategy's edge is more fragile.

Measured spreads were 2-3x wider than normal trading conditions during earnings windows: 2.8% median round-trip cost for Micro-Cap, 1.6% for Small-Cap, and 0.7% for Mid-Cap. After these costs, the High-Hype strategy's 8.17% gross return falls to 6.57% net (80% of alpha survives), while the ML strategy's 4.07% gross falls to just 2.17% net (53% survives)—still positive, but far more fragile to execution quality. The high-hype strategy's concentration in higher-quality, lower-spread Small-Cap opportunities makes it the more economically robust choice for real-world implementation.

Implications for Market Efficiency

These findings do not overturn the Efficient Market Hypothesis so much as define its boundaries: markets are mostly efficient (the best model achieves only 55.8% accuracy), but efficiency is a spectrum, and small-cap earnings with elevated attention show inefficiencies that persist long enough (5 days) to exploit. This persistence likely reflects capacity constraints (the $50-100 million total addressable market before price impact dominates is trivial for hedge funds but meaningful for retail), career risk (a strategy with 5-day holding periods and lumpy, quarter-to-quarter volatility is risky for professional managers to run), and structural friction (daily monitoring of 645 stocks is too labor-intensive for institutions to automate profitably at this scale but is feasible for a disciplined individual trader).

The Self-Defeating Prophecy

A natural question for any published trading strategy is whether publicizing it destroys the edge. The honest answer is that it would likely degrade but not disappear. If many traders began screening for hype ≥ 7.0, competition would compress the pattern—bidding up prices pre-earnings, worsening execution, and pushing sophisticated competitors toward front-running the threshold itself. But the underlying driver, retail investors' behavioral over-reaction to

pre-earnings momentum, is a cognitive bias, not an information gap, and such biases do not disappear just because they are documented: PEAD itself has persisted for over 50 years despite being well known, compressing from roughly 4-5% in the 1970s (Bernard and Thomas) to 2-3% today rather than vanishing. A reasonable projection is a 3-5 year half-life for the high-hype edge—intact in years 1-2 as only early, small-scale adopters implement it, moderately degraded (6-7% → 4-5%) in years 3-4 as more sophisticated players automate it, and stabilizing around 3-4% by year 5 as the strategy reaches equilibrium between behavioral persistence and competitive pressure. That outcome is still meaningful, and it is also the expected and appropriate result of academic research: documenting inefficiencies should push markets toward greater efficiency over time, not preserve alpha indefinitely.

Contributions to the Literature

This research contributes in three ways: it shows that a data-driven, empirically optimized composite metric (the 20/60/20 hype score) outperforms an intuition-based one by 72%; it demonstrates that machine learning adds value in finance primarily by capturing non-linear interactions rather than through black-box complexity, since the Random Forest improvement over logistic regression, while numerically modest (2.8 percentage points), is statistically robust and economically meaningful at scale; and it offers a fully reproducible event-study methodology that balances theoretical grounding with empirical flexibility. Most broadly, it shows that market inefficiencies in small-cap earnings are not just academic curiosities but a genuine, if modest, edge accessible to informed retail traders willing to do the work—precisely because behavioral factors dominate in a niche too small and labor-intensive for institutional capital to fully arbitrage away.

Limitations

Several limitations should be noted. Survivorship bias means the dataset only includes stocks trading as of 2024, excluding failures and acquisitions, which likely inflates returns somewhat. While transaction costs were measured empirically, the analysis does not model market impact from larger position sizes or slippage from rapid post-earnings price moves. Look-ahead bias exists because hype score weights were optimized on the full dataset rather than a separate training period. The high-hype sample is limited to 35 events, creating wide confidence intervals despite the robustness checks performed. The 2022-2024 sample window means market-regime dependence is untested, no shorting or hedging strategies were explored, and price-efficiency assumptions may not hold around gaps, halts, and illiquid closes.

Conclusion and Future Research

This study provides robust evidence that pre-earnings market signals predict post-earnings drift in small-cap stocks. The hype score (60% momentum, 20% volume, 20% coverage) identifies high-attention events generating 260% outperformance relative to baseline, a result that survives outlier removal, transaction cost analysis, and sample-size scrutiny. The

$300M-$1B market-cap range is a genuine sweet spot balancing liquidity and persistent inefficiency, and the zero-coverage paradox shows that maximum information asymmetry does not translate into maximum opportunity once illiquidity and lack of validation are accounted for. Machine learning enhances but does not replace simple threshold rules: Random Forest's 55.8% accuracy captures real non-linear structure, but the simple hype ≥ 7 filter's higher selectivity produces superior risk-adjusted and post-cost returns, suggesting that in low signal-to-noise domains like financial markets, interpretable rules grounded in theory can outperform more complex models. Future work should prioritize genuine out-of-sample testing (training on 2022-2023, testing on 2024 alone), more granular market-impact and slippage modeling, sector-level analysis, and short-side/market-neutral variants. Promising extensions include incorporating options-market signals (implied volatility, put/call ratios) and social-media sentiment, expanding the sample period back to 2015 to test across market regimes, and eventually validating the strategy through live paper trading. For practitioners, the recommended implementation is straightforward: focus on the Small-Cap tier ($300M-$1B), use hype score ≥ 7 as the primary filter, supplement with ML 50%-confidence signals, require positive relative strength versus the S&P 500, size positions by volatility, and hold for 5 trading days. While publication may erode part of this edge over time, the underlying behavioral and structural mechanisms suggest it will not disappear entirely, and even a 50% degradation over five years would remain economically meaningful for disciplined retail traders.

Acknowledgements

I would like to thank my mentor, Vijay Krishnan, for guidance throughout this research project and for encouraging rigorous statistical validation of findings. I am grateful to the open-source data science community, particularly the developers of pandas, yfinance, scikit-learn, and matplotlib, whose tools made this research possible without expensive data subscriptions. I acknowledge the academic researchers whose prior work on PEAD (Ball and Brown; Bernard and Thomas), momentum (Jegadeesh and Titman), and attention effects (Barber and Odean) provided the theoretical foundation for this study. Finally, I thank my family and friends for their patience during the many hours spent coding, analyzing data, and writing this paper.

Works Cited

Ball, Ray, and Philip Brown. “An Empirical Evaluation of Accounting Income Numbers.” Journal of Accounting Research, vol. 6, no. 2, 1968, pp. 159–178.

Barber, Brad M., and Terrance Odean. “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors.” The Review of Financial Studies, vol. 21, no. 2, 2008, pp. 785–818.

Bernard, Victor L., and Jacob K. Thomas. “Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?” Journal of Accounting Research, vol. 27, 1989, pp. 1–36.

Breiman, Leo. “Random Forests.” Machine Learning, vol. 45, no. 1, 2001, pp. 5–32.

Chordia, Tarun, and Lakshmanan Shivakumar. “Earnings and Price Momentum.” Journal of Financial Economics, vol. 80, no. 3, 2006, pp. 627–656.

Cohen, Jacob. Statistical Power Analysis for the Behavioral Sciences. 2nd ed., Lawrence Erlbaum Associates, 1988.

“Complete List of NASDAQ and NYSE-Listed Securities.” NASDAQ Trader FTP Service, Nasdaq, accessed Jan. 2025.

Fama, Eugene F. “Efficient Capital Markets: A Review of Theory and Empirical Work.” The Journal of Finance, vol. 25, no. 2, 1970, pp. 383–417.

Jegadeesh, Narasimhan, and Sheridan Titman. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” The Journal of Finance, vol. 48, no. 1, 1993, pp. 65–91.

McKinney, Wes. “Data Structures for Statistical Computing in Python.” Proceedings of the 9th Python in Science Conference, 2010, pp. 56–61.

Murphy, Kevin P. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.

Pedregosa, Fabian, et al. “Scikit-Learn: Machine Learning in Python.” Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.

Alkali Metal Anodes for Rechargeable Batteries By Tim Zeng

Abstract

The growing demand for low-carbon energy technologies has increased the need for rechargeable batteries with higher energy density, longer cycle life, and improved safety. Among the key components of rechargeable batteries, the anode plays a central role in determining overall capacity, efficiency, and commercial feasibility. While electricity can be a viable option, there are certain advancements that need to be made, and there are also obstacles that impede the advancements in batteries. This review examines the potential of alkali metal anodes, including lithium, sodium, and potassium metal anodes, as alternatives to conventional graphite-based anodes. It compares their theoretical advantages in energy density and cost with the major barriers to practical application, including dendrite growth, unstable solid electrolyte interphase formation, low Coulombic efficiency, and safety risks. The review further evaluates emerging strategies, such as artificial SEI layers, surface modification, electrolyte engineering, and three-dimensional host structures, to assess whether alkali metal anodes could become commercially viable in the coming decades.

Development and Feasibility of Alkali Metal Anodes

Historical Development of Battery Anodes

The development of batteries can be traced back more than two centuries to Alessandro Volta's voltaic pile, an early electrochemical device that demonstrated the possibility of producing a continuous electric current through chemical reactions (Ming et al.). This invention established the basic principle of using two electrodes and an electrolyte, which remains central to modern battery design. A century later, the first rechargeable battery was developed. The rechargeable, or secondary batteries have been constantly improved in the last century, and this leads us to discussion of alkali anodes (Ming et al.). Although lithium ions are widely used as charge carriers in commercial lithium-ion batteries, lithium metal itself is not commonly used as the anode in current mainstream cells. Early interest in lithium metal anodes was limited by safety concerns, especially dendrite formation and unstable interfacial reactions (Kaliaperumal et al.). As a result, graphite became the dominant anode material because it offers better cycling stability and safer lithium intercalation. However, this improvement in safety and reliability comes at the cost of a lower theoretical capacity compared with metallic lithium (Seroka et al.).

Current Development of Alkali Metal Anodes

Currently, the demands for cleaner energy sources than fossil fuels highlight the necessity for efficient rechargeable batteries, a cleaner energy source than fossil fuels for the energy transition. The urgent need for massive amounts of energy creates challenges for many different aspects of rechargeable batteries, such as energy density, storage capacity, sustainability, and recyclability (Ngoy et al.). One promising direction is the development of "post-lithium" battery

systems, including sodium- and potassium-based batteries, as well as renewed interest in lithium metal batteries. These systems are attractive because alkali metals can offer high theoretical capacities and low redox potentials. Sodium and potassium are also more abundant than lithium, which may reduce material cost and supply-chain pressure (Tedla et al.). However, the benefits vary significantly among Li, Na, and K metal anodes, so their technical and economic feasibility should be evaluated separately.. Conventional lithium-ion batteries are limited by the theoretical capacities of their intercalation-based electrode materials (Idu et al.). In contrast, metal-anode systems, such as lithium metal, sodium metal, and sodium–oxygen batteries, can offer much higher theoretical energy densities. However, these values are usually calculated at the cell-chemistry level rather than for the anode alone, so direct comparisons must clearly define whether the value refers to the anode material, the full cell, or the practical battery pack. Another reason the metal anode is so attractive is the development of alkali metal free cathodes, which possess high capacity due to multi-ion reactions but requires pairing up with an alkali metal containing anode to function since it still uses alkali metal ions for transporting the charges (Mashfy et al.). Options for alkali metal anodes mainly include Li, Na, and K. While Li is the most promising choice with its highest energy density (3860 mAhr g-1) (Ngoy et al.; Idu et al.).

Economic Feasibility and Cost Considerations

The economic feasibility of alkali metal anodes should be evaluated using several separate metrics, including raw material cost, processing cost, capacity-normalised cost, and full-cell manufacturing cost (Mashfy et al.). Although lithium-ion battery pack prices have decreased substantially over the past decade, replacing graphite with a metal anode does not automatically reduce total cost, because additional requirements such as protective layers, modified electrolytes, pressure control, and safety management may increase manufacturing complexity (X. Li et al.). Therefore, the cost advantage of alkali metal anodes should be discussed together with their expected improvement in specific capacity and energy density. M anodes (36 $ kA/h) are still cheaper compared to the traditional material such as graphite (70 $ kA/h), which means adapting to alkali metal (Xiang et al.).

Key Challenges for Alkali Metal Anodes

These challenges mainly arise from the high chemical reactivity of alkali metals and their unstable behaviour during repeated plating and stripping. Uneven metal deposition can lead to dendrite growth, which may pierce the separator and cause internal short circuits. At the same time, repeated SEI fracture and regeneration consume active metal and electrolyte, reducing Coulombic efficiency and shortening cycle life (He et al.). These electrochemical problems increase safety risks and may also raise production and maintenance costs. These problems not only posed a threat to safety, but also affected the practicality of this technology through raising production and maintenance costs. They prohibited the full-scale commercialization of alkali metal anodes in rechargeable batteries, even if it brings significant boost in performances (Idu et al.; Chen et al.).

Technological Barriers and Proposed Solutions

Safety Barriers and Dendrite Formation

Due to the immense potential of alkali metal anodes, scientists had proposed numerous solutions to accelerate the application of it. However, several technological barriers continue to limit the practical application of alkali metal anodes, particularly safety risks, unstable interfacial chemistry, poor cycling stability, and the difficulty of maintaining uniform metal deposition under practical operating conditions. To begin with, there are some issues that directly impact the safety of the battery, making the alkali metal anode a risky choice (Xiang et al.). These problems are primarily caused by the high reactivity of the alkali metals, leading to the unstable nature of the anode that can be disastrous. During repeated plating and stripping, metal ions are reduced and deposited on the anode surface. If the local current density, ion concentration, or surface morphology is uneven, metal deposition becomes non-uniform. Small protrusions on the anode surface can concentrate the electric field and attract further ion deposition, gradually developing into dendritic structures. These metallic peaks pose a severe threat to the battery since it not only affects the normal function, but also might penetrate the separator between the cathode and the anode which directly causes short circuit, potentially burning and even explosion (Wang et al.; Xiang et al.). This requires measures to either suppress metal uneven deposition or enhance the separator's hardness to reach complete safety, and will be discussed later. Other than safety concerns, there are also issues that impact efficiency, making it less profitable when applying the technology. On the surface of the metal anode, a thin passivation layer known as the solid electrolyte interphase, or SEI, forms through side reactions between the highly reactive metal and the electrolyte. Ideally, the SEI should allow alkali metal ions to pass through while blocking electrons, thereby reducing continuous electrolyte decomposition and stabilising the electrode–electrolyte interface (Clare et al.).

Solid Electrolyte Interphase Instability

However, the naturally formed layer is thin, thermodynamically unstable and therefore fragile. If it is fractured, the metal anode further reacts with the electrolyte (undesirable because it consumes the charge transferring ions and the electrolytes) and causes a drop in the Coulombic efficiency. The layer then formed after the fracture tends to be thicker and has a significant impact on the transfer of ions. Overall, this important but fragile structure needs careful treatments, and directly affects the output level, which is the main advantage of the alkali metal anodes (N. W. Li et al.).

Surface Modification and Artificial SEI Strategies

To respond to these issues, some futuristic but achievable proposals had been brought. Starting with the metal dendrite deposition, scientists had tried to figure out the complete mechanism of dendrite forming since the 1980s, but it was still not fully understood. However, there is an increasing focus on this topic and therefore many solutions are proposed. The uneven

distribution of the alkali metals generally emerges on the foils of alkali metal anodes where topographical and electrical inhomogeneities are present. Such defects can be decreased by removing the naturally formed layer on the alkali metal foil, leading to higher uniformity of alkali metal deposition and therefore hinders the growth of dendrites (Dachraoui et al.). Moreover, low current densities also helped to counteract with the formation of the dendrites, and a larger surface area of the anode foil would effectively decrease the current density.* To achieve this, micro and/or nano structuring of alkali metal anodes were considered as a viable method and experiments were conducted, and Li foils with surface modification have a much higher surface area and therefore effectively decreased the dendrite activity.* However, due to direct exposure to electrolytes, the surface micro structures cannot last a long time before becoming defunct, and therefore is a question requiring increased effort. For the efficiency and safety related SEI layer, there are also some proposed solutions (Ningappa et al.). The issues and concerns regarding this structure mainly comes from its fragile nature, which impacts the Coulombic Efficiency and shortens the lifespan (through consuming active materials such as the alkali metal and electrolyte when restoring itself). To address this issue, researchers have explored artificial SEI layers formed through methods such as chemical pre-treatment, polymer coating, inorganic protective layers, and electrolyte additive design. An effective artificial SEI should combine high ionic conductivity with low electronic conductivity, strong mechanical stability, and chemical compatibility with the metal anode. Such a layer can regulate ion flux, reduce direct contact between the metal and electrolyte, and promote more uniform metal deposition beneath the protective interface (Alam et al.; Yang et al.).

Summary and Commercialisation Outlook

In summary, the pursuit of alkali metal anodes represents a critical frontier in the quest for next-generation energy storage, directly addressing the global energy crisis driven by the depletion of fossil fuels. The core issue lies in the inherent instability of these anodes, where high reactivity leads to dendrite growth that threatens safety through short-circuits, and a fragile solid electrolyte interface that continuously fractures and reforms, consuming active materials and crippling Coulombic efficiency—undermining the very energy density advantage that makes them so attractive (Ngoy et al.; Mashfy et al.). The proposed solutions target these weaknesses with direct interventions: dendrite suppression is approached by engineering the anode's physical architecture through micro- and nano-structuring of the metal foil, which increases surface area to lower local current density and smooths out topographical defects that seed uneven deposition. Meanwhile, the fragile natural SEI layer can be bypassed entirely by constructing an artificial, ex-situ layer, one meticulously designed to be mechanically robust yet ionically conductive, guiding uniform metal deposition beneath it and preventing ongoing electrolyte decomposition (Fu et al.). While scientifically sound in controlled settings, the realism of these solutions for large-scale application is tempered by unresolved practicalities; surface structuring degrades rapidly under prolonged electrolyte exposure, and artificially perfect SEI layers face the immense challenge of maintaining their integrity across thousands of charge-discharge cycles in

a cost-effective manufacturing process. This is not merely a laboratory curiosity—without a radical improvement in how we store and dispatch intermittent renewable energy; the transition away from non-renewable resources will stall. Alkali metal batteries, with their potential for unmatched energy density and use of abundant materials like sodium, offer a tangible path to making grid-scale storage and long-range electrified transport economically viable, thereby striking directly at the heart of the energy dilemma (He et al.; Mashfy et al.; Ngoy et al.).

Works Cited

Alam, Saud, et al. “Interface Chemistry in Solid-State Batteries: Advances, Challenges, and Future Prospects.” Results in Chemistry, vol. 18, Nov. 2025, p. 102832. https://doi.org/10.1016/j.rechem.2025.102832.

Chen, Siyan, et al. “Safety Challenges and Safety Measures of Li-Ion Batteries.” Energy Science & Engineering, vol. 9, no. 9, 2021, pp. 1647–72. https://doi.org/10.1002/ese3.895.

Clare, A. T., et al. “Metal Multi-Material Additive Manufacturing: Overcoming Barriers to Implementation.” CIRP Annals, vol. 74, no. 2, Jan. 2025, pp. 869–93. https://doi.org/10.1016/j.cirp.2025.05.004.

Dachraoui, Walid, et al. “Nucleation, Growth and Dissolution of Li Metal Dendrites and the Formation of Dead Li in Li-Ion Batteries Investigated by Operando Electrochemical Liquid Cell Scanning Transmission Electron Microscopy.” Nano Energy, vol. 130, Nov. 2024, p. 110086. https://doi.org/10.1016/j.nanoen.2024.110086.

Fu, Yonghuan, et al. “Interface Engineering in Solid-State Sodium Batteries: Advances in Materials Design, Interfacial Strategies, and Multiscale Characterization.” Interdisciplinary Materials, vol. 5, no. 1, 2026, pp. 5–84. https://doi.org/10.1002/idm2.70023.

He, Pan, et al. “Toward Zero-Excess Alkali Metal Batteries: Bridging Experimental and Computational Insights.” Advanced Materials, vol. 37, no. 46, 2025, p. 2502052. https://doi.org/10.1002/adma.202502052.

Idu, Cyril Ikechukwu, et al. “A Review of Recent Advances, Current Limitations, and Remedies of Lithium-Ion Batteries for Advanced Technological Applications.” Future Batteries, vol. 7, Sept. 2025, p. 100109. https://doi.org/10.1016/j.fub.2025.100109.

Kaliaperumal, Muthukrishnan, et al. “Cause and Mitigation of Lithium-Ion Battery Failure—A Review.” Materials, vol. 14, no. 19, Jan. 2021, p. 5676. https://doi.org/10.3390/ma14195676.

Li, Nian Wu, et al. “Passivation of Lithium Metal Anode via Hybrid Ionic Liquid Electrolyte toward Stable Li Plating/Stripping.” Advanced Science, vol. 4, no. 2, 2017, p. 1600400. https://doi.org/10.1002/advs.201600400.

Li, Xueqian, et al. “Reutilization and Upcycling of Spent Graphite for Sustainable Lithium-Ion Batteries: Progress and Perspectives.” eScience, vol. 5, no. 4, July 2025, p. 100394. https://doi.org/10.1016/j.esci.2025.100394.

Mashfy, Mohammad Muhtasim, et al. “Sodium Ion Batteries: A Sustainable Alternative to Lithium-Ion Batteries with an Overview of Market Trends, Recycling, and Battery Chemistry.” Next Energy, vol. 10, Jan. 2026, p. 100478. https://doi.org/10.1016/j.nxener.2025.100478.

Ming, Fangwang, et al. “Perspective on Aqueous Batteries: Historical Milestones and Modern Revival.” Advanced Materials, vol. 38, no. 14, 2026, p. e72294. https://doi.org/10.1002/adma.72294.

Ngoy, Kitalu Ricin, et al. “Lithium-Ion Batteries and the Future of Sustainable Energy: A Comprehensive Review.” Renewable and Sustainable Energy Reviews, vol. 223, Nov. 2025, p. 115971. https://doi.org/10.1016/j.rser.2025.115971.

Ningappa, Ningaraju Gejjiganahalli, et al. “Mechanistic Insights and Materials Strategies for Dendrite-Free Metal Anodes in Alkali and Zinc Batteries.” Nano Energy, vol. 141, Aug. 2025, p. 111144. https://doi.org/10.1016/j.nanoen.2025.111144.

Seroka, Ntalane Sello, et al. “Recent Advancement and Prospects of Graphite Nanocomposites as Anode Materials for Lithium-Ion Batteries.” Renewable and Sustainable Energy Reviews, vol. 238, Sept. 2026, p. 117066. https://doi.org/10.1016/j.rser.2026.117066.

Tedla, Tsegaye Sissay, et al. “Advancements in Sodium-Ion Batteries Technology: A Comprehensive Review of Recent Development on Materials, Mechanisms, Applications, and Prospects for Energy Storage.” Energy Reports, vol. 14, Dec. 2025, pp. 3175–203. https://doi.org/10.1016/j.egyr.2025.10.014.

Wang, Hua, et al. “Alkali Metal Anodes for Rechargeable Batteries.” Chem, vol. 5, no. 2, Feb. 2019, pp. 313–38. https://doi.org/10.1016/j.chempr.2018.11.005.

Xiang, Jingwei, et al. “Alkali-Metal Anodes: From Lab to Market.” Joule, vol. 3, no. 10, Oct. 2019, pp. 2334–63. https://doi.org/10.1016/j.joule.2019.07.027.

Yang, Yunlong, et al. “Electrolyte Engineering toward Rational Electrode–Electrolyte Interfacial Designs for Metal Batteries.” Advanced Science, vol. 12, no. 48, 2025, p. e16870. https://doi.org/10.1002/advs.202516870.

Climate Governance in Denmark and Australia: A Comparative Analysis and Lessons for Climate Action By Elizabeth Townsley

Abstract

Climate Change remains one of the most pressing issues of our time, requiring fast-paced and adaptive responses to mitigate its far-reaching impacts. This paper aims to explore actions and legal measures undertaken to limit climate change, particularly the Paris Climate Agreement, whilst also discussing legal landmark cases around the issue. In particular, a focus on Denmark, a leading country in climate mitigation, and on Australia, a country with multiple obstacles preventing comprehensive climate mitigation, is established through a comparative study. These two jurisdictions present contrasting models of climate governance, providing valuable insights into how differing legal frameworks, economic structures, and policy approaches influence the implementation of international climate commitments. Both countries' national obligations and climate policies are analysed, with references to the successes and shortcomings of each country's strategic roadmap. In conclusion, the paper contrasts the two strategies broadly in an attempt to establish key lessons and best practices from both countries to better inform effective climate governance and strengthen global climate resilience.

Introduction

Climate change is the greatest challenge to the advancement and progress of mankind, affecting people from all parts of the world. The key to addressing this challenge lies in focused and timely interventions, linking local, national and international efforts. This has been shown in the form of many international treaties and conventions which have been ratified in most countries, either federally, on a state level, or implemented at the local level, within communities.

In this regard, one of the most influential conventions is the Paris Agreement which was introduced in 2016, and now has 194 parties (including 193 states and the European Union). The main objective of the Agreement is to limit global temperature rise in this century to well-below 2 degrees Celsius, while making efforts to limit it to 1.5 degree celsius (pre-industrial levels). Studies suggest that in order to achieve the Paris climate goals, the global community must triple renewable energy usage by 2030. To ensure and monitor compliance, Global Stocktakes are taken every five years, which include each state's Nationally Determined Contribution (NDC), outlining their updated national climate plan in the short-term to reduce Greenhouse Gas (GHG) Emissions, and the degradation of the ozone layer, either by assisting developing countries in achieving their climate goals, or by implementing federal laws, encouraging local communities to take more action, whilst also having a legislated goal to hold the government accountable if those goals are not reached. It can also include optional long-term goals of the country. The Committee to Facilitate Implementation and Promote Compliance consists of 12 members, each serving for a term of three years, and it assists with the implementation of the Paris Agreement.

Although there is no direct punishment for not achieving these goals, it is in each state's best interests to achieve them.

However, the Paris Agreement is not the only major convention. It covers climate change on a more general scale, whilst other treaties focus on more specialised areas of climate change. Some of these are the Montreal Protocol on Substances that Deplete the Ozone Layer, 1987 (MPSDOL), the Convention on Biological Diversity, 1992 (CBD), and the United Nations Convention to Combat Desertification, 1994 (UNCCD). These are each designed to target specific aspects of climate change, with the MPSDOL regulating the production and consumption of nearly 100 ozone depleting chemicals, the CBD promoting actions with a goal of leading to a more sustainable future, and the UNCCD addressing desertification and the effects of droughts. These have all been ratified by either 196 or 197 parties, and are extremely important in reaching the ultimate goal of net zero emissions, which is defined as reaching a state in which climate pollution going into the atmosphere is balanced by removal out of the atmosphere.

At present, Denmark is one of the top countries in climate action in the world, with societal attitudes contributing significantly towards the shift to a greener future. Currently, Denmark's Climate Change Act, 2020 (CCA) has legislated many goals which are similar to the Paris Agreement goals. Meanwhile, Australia, due to various factors such as main industries and exports, is currently not reaching its federal legislation in regards to climate change, meaning that there is an inconsistency towards Australia fulfilling its international obligations regarding climate change.

In this paper, the author will attempt to address the national obligations of Denmark and Australia in relation to climate change, and summarise their respective advantages and disadvantages in terms of action taken to limit the effects of climate change, and conclude with an analysis to determine global best practices.

Legal Background: The Evolution of Climate Governance

Climate change has been a prominent issue globally, becoming major in the mid-1800s, with issues such as large levels of carbon dioxide, and the reduction of biodiversity being major consequences of the actions of humans globally. In 1911, atmospheric carbon dioxide levels were at the highest it had been in any point in the Earth's history, and currently, carbon dioxide levels from human activity have been increasing around 250 times faster than it did from natural sources after the last ice age. Realising and understanding these changes have been the basis to most conventions, treaties and principles established to limit these activities.

As mentioned above, the period in which climate change was recognised as a global issue and the effects were becoming increasingly obvious, many treaties and conventions were introduced to try and mitigate these effects and limit the damages caused. The majority of them were created from 1987 - 1994, and targeted more specific areas of climate change, such as biodiversity, or harmful chemicals which, at the time, were being mass produced. During this time, principles were also made and adopted into various conventions to outline ways in which

all countries are able to achieve their goals no matter their financial capacity or current emissions.

A major principle is that of Common But Differentiated Responsibilities (CBDR) which is a key example of a principle that has been widely adopted into different treaties, and has been recognised globally. CBDR is a principle which was established by the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 at the first Earth summit in Rio de Janeiro. Since then, it has been adopted into many international treaties, and was incorporated into Section 2 of the Paris Agreement. The principle focuses on financially assisting areas that are severely affected by climate impacts but are economically disadvantaged. This focuses on countries which are more industrially focussed, so produce more emissions such as Japan, India, and China. Whilst not all manufacturing countries are considered to be developing countries, the majority are, and do not have the finances to implement strategies to offset their carbon dioxide emissions. This principle ensures that more developed countries, who have less emissions due to their main industries, such as Sweden, the UK, and Italy, provide financial assistance to those developing countries to ensure that their carbon dioxide emissions are limited as much as possible.

Climate litigation is the act of fighting for climate justice in a legal environment. This form of advocacy is in place all around the world, with the leading countries being the USA and Australia, presumably due to the exacerbated effects being experienced by these countries, such as more and fiercer wildfires, and decreased biodiversity. There have been many landmark cases around the world, which are holding individuals, and the government accountable for failing to limit the effects of climate change occurring. An example of this can be seen in Australia's first successful case surrounding climate change, the Gloucester Resources Limited v. Minister for Planning (2017) case. Here, a coal mine was denied being built by the Australian government due to the project not being in the public's best interests. This was due to the direct link between the project and carbon dioxide and other emissions and climate change. It was proved that the costs of that coal mine outweighed the benefits it would have on society economically, socially, and especially, environmentally. There have been many similar cases around the world, each fighting for other aspects of climate change such as the USA: Massachusetts v. Environmental Protection Agency (2007) case which shows the ratification of the MPSDOL. The case called for the Environmental Protection Agency (EPA) to begin regulating the emissions of four ozone depleting gasses, including carbon dioxide. This area of law is slowly adjusting to current society's situation and is integral in ensuring that ratified treaties and conventions are upheld in order to mitigate climate change and its effects on the environment, and the world.

Denmark: Setting the Benchmark for Climate Action

Denmark is a leading country in the mitigation of climate change, with its current laws and strategies placing it first in the world. This can be partly attributed to the agrarian economy, the principal exports mainly being medical supplies such as vaccines or hormones, and their natural resources being limited, leading to a relatively small sector in mining. Another factor is

the public perception of climate change, with nearly 74% of the population seeing it as one of the four most serious issues to be dealt with, both by the government and citizens.
Denmark has incorporated a National Energy and Climate Plan for the period 2021-2030, which outlines their plan to combat climate change. This takes into account the Paris Agreement target, and aims to reduce Danish greenhouse gas emissions by 70 % by 2030 compared to 1990 levels, achieve climate neutrality by 2045 at the latest, and have a 110% reduction in emissions by 2050. Some of the ways in which the country is planning on achieving these goals is by phasing out petrol and diesel cars by 2030, and expanding renewable energy usage, particularly by introducing offshore wind farms in the North Sea.

Further, the Danish Council on Climate Change consists of one chair and eight members, and provides recommendations to the Minister for Climate, Energy and Utilities in making decisions regarding climate change. Members have a four year term, and can be reappointed by the Danish Minister of Climate and Energy once. They are each experts with a broad expertise and a high level of climate relevant academic knowledge. The council is assisted by a secretariat which consists of one secretary and twenty full time employees. The council assesses whether the government's climate efforts make it probable that the country's climate targets will be reached, and provides a status update on Denmark's international targets. There are four key targets, which are to 1) evaluate the status of Denmark's implementation of national climate objectives and international climate commitments, 2) analyse potential means of transitioning to a low-carbon society by 2050 and identify possible measures to achieve greenhouse gas reductions, 3) draw up recommendations to help shape climate policy, including a selection of potential mechanisms and transition scenarios, and 4) contribute to the public debate. The Danish Council on Climate Change must, to the extent required in the preparation of its analyses and other work, consult and involve relevant parties, including, among other business interests, social partners in the labour market and civil society.

To ensure that this plan works, the Danish government is prepared to financially support these goals with annual funding, meaning that they could be achieved through subsidies alone. The budget of the government was set to gradually increase until it reached the Danish Kroner (DKK) amount of 2.8 billion in 2025, which is equivalent to 0.1% of the GDP of Denmark. Since the start of the Climate Act, about 120 billion kroner has been allocated to the transition, and approximately 5 billion more will be allocated by 2030 to further efforts for the environment. After some financial agreements were made on April 15th 2024 regarding more funding for the Act, the Minister for Climate, Energy, and Utilities, Lars Aagaard commented,"To achieve the green transition, we need to make the green choice easier. We do that with today's agreement."

Denmark is leading the way globally for climate change, introducing legislation and a legislated plan to combat climate change and achieve net zero. This has been financially supported by the government, and a committee has been introduced to oversee the targets, and ensure they are being reached. Denmark is setting an example to other countries of what legislated action looks like, and the progress it achieves.

Australia: The Challenge of Translating Ambition into Action

Even as most Australians (nearly 70%) think that climate change is a prevalent issue, Australia's economy remains heavily dependent on resource exports. Australia's main exports are primarily concerned with the large mining industry, including iron ore, coal, and natural gas.

Australia has committed to climate goals, as legislated in the Climate Change Act, 2022. Despite this, and due to current issues in society, these goals are currently not being reached, and further international issues have affected Australia's ability to achieve its goals under Act. Australia has incorporated its Climate Change Act into everyday actions, aiming to achieve 43% reduction targets in greenhouse gas emissions from 2005 levels by 2023 and net zero by 2050. It has also promised a $4.6 billion spending commitment towards climate action. To ensure that these goals are reached, the Climate Change Authority was established under the Climate Change Authority Act, 2011. The Authority functions with the sole purpose of providing expert advice to the Australian government to achieve climate targets and lead Australia towards a prosperous, resilient and decarbonised future. In the 2025 annual report, it stated that the rate of reductions must double in the 5 years to 2030 and triple in the decade to 2035 in order to stay on track with the stated goals. In order to achieve this, the authority has provided a list of seven recommendations for priority action which must be considered to have the best chance of achieving the Act's long–term emission targets.

Australia has built on these seven recommendations by identifying five priority areas and expanding on each to help and keep the country on track. These recommendations are: Clean electricity across the economy, Lowering emissions by electrification and efficiency, Expanding clean fuel use, Accelerating new technologies, and, Net carbon removals being scaled up. The main focus of these points is to reduce greenhouse gas emissions enough to meet the legislated requirements. They focus on the general household, electric cars, green electricity, the discovery of greener technologies, and supporting land owners by setting up an Australian Carbon Credit Unit (ACCU) scheme.

Despite these efforts, fossil fuels are still a large part of the Australian economy, leading to many fossil fuel subsidies, with the figure reaching $16.3 billion in 2026. They have been introduced due to the high use and reliance of the economy on fossil fuels, and it provides the public with reduced/subsidised costs for fossil fuels. These subsidies involve refunding mining companies, farmers, and other industries, the excise paid on petrol and diesel. The critics of these subsidies say that refunds encourage greater burning of fossil fuels and work against policies that incentivise businesses to embrace lower emissions vehicles and other technology. There are calls from such groups for the subsidies to be cut back, with Rob Campbell, the Australian institute's research director saying, "Cutting back subsidies like these, which make the community and the climate worse off, are an obvious priority for any government that says it is concerned with the budget bottom line, inequality, or climate change."

To make these goals more difficult to achieve, the 2025 Advisory Opinion on Obligations of States in Respect of Climate Climate by the International Court of Justice ("ICJ") found that Australia was liable for emissions due to the exportation of any material from the coal and gas

industry. However, the Australian government has disagreed with this ruling, stating that the Paris Agreement means Australia is not responsible for the harm caused to the climate system by GHG emissions from fossil fuels exported overseas, and that the incremental emissions from each new coal mine or gas facility in Australia are too small to constitute a significant impact on the global climate or Australia's environment. These statements by Australia have been used in extending coal and gas projects despite them being inconsistent with the ICJ ruling. Australia has still not taken any steps in the mitigation of its emissions in relation to this ruling, causing the net zero goals to be unachievable according to the ICJ.

Although Australia has taken legislative steps to try and ensure that various climate goals are going to be reached, such as net zero by 2050, it remains to be seen how such ambitious targets can be met at the current rate of emissions. Not only is Australia providing fossil fuel subsidies, which are contributing to Australia's GHG emissions, but it is also disagreeing with the ICJ ruling on fossil fuel emissions from exported fossil fuels contributing to Australia's GHG emissions. At the current rate of mitigation steps, Australia will not reach its legislated climate goals, and thus, it is imperative that Australia improve its mitigation strategy and implementation for net zero to be reached.

Comparative Analysis: Lessons from Two Climate Governance Models

Denmark and Australia are two countries with contrasting approaches in regards to climate change goals and their attitudes towards climate change. Both have created NDCs and a legislated goal to combat climate change, however, both countries also have different methods of implementation, and financial goals with regards to how these legislations will succeed.
Both countries have passed dedicated legislations for climate change, through Australia's Climate Change Act (2022), and Denmark's National Energy and Climate Plan (2019) for the period 2021-2030. Denmark has set somewhat more ambitious targets, with its aim to reach climate neutrality by the latest 2045, and have a 110% reduction in greenhouse gas emissions by 2050, compared to Australia's climate neutrality by 2050. Despite these slight differences each country has made goals which they believe are most achievable to them due to their circumstances. Australia, due to their current circumstances, has set less ambitious climate goals. This is partly due to a large sector in mining that Australia has, relying on those exports to keep its economy running, meaning that it is practically difficult for Australia to reach net zero. Denmark, on the other hand, has set extremely ambitious goals, which it is confident that it will achieve. This is due to their extremely secure programs, such as expanding renewable energies and phasing out diesel and petrol cars, which are already in place, with a reduction in GHG emissions already being seen.

Overall, Denmark has taken more steps in the reduction of GHG emissions, causing the country to receive an overall higher climate combatting ranking compared to Australia across multiple international indices on climate action. For example, the Climate Change Performance Index (CCPI) of Denmark is currently, as of 2026, ranked 4th in the world for climate action, receiving a 'high' ranking, with an overall score of 80.5. In contrast, Australia is ranked 56th,

with a ‘very low’ rating, and an overall score of 42.5. This can also be seen as in the Sustainable Development Report 2025 published by the Dublin University Press, where Denmark was ranked third in the world for upholding the Sustainable Development Goals (SDGs), with a score of 85.3, compared to Australia being ranked 36, with a score of 77.9.

Both Denmark and Australia have also collaborated to set climate goals, as seen in the 2023 NATO summit, where the Prime Minister of each state at the time signed a Strategic Partnership Arrangement in Vilnius. This was the point at which it was decided that both countries would limit temperature increase to 1.5 degrees Celsius above pre-industrial levels, as set out in the Paris Agreement. During March 2026, the Danish had an official state visit to Australia, represented by Their Majesties King Frederik X and Queen Mary of Denmark, which focussed on agri-food. This is because both countries share a passion in agriculture and food production, and stated their desire to collaborate with some new innovations to achieve more sustainable farming and agricultural practices.

Although both Denmark and Australia have taken respective actions to limit climate change, and have worked together on prominent issues and goals, Denmark has been more effective in addressing climate change. This can be seen as Denmark is ranked more highly compared to Australia in multiple reports, including the CCPI, and the Sustainable Development Report of 2025. These results convey how Denmark’s climate actions have been more effective in mitigating climate change compared to Australia, and therefore showing the way forward for countries who are having difficulty combating climate change.

Conclusion: Charting the Way Forward

Climate change is an extremely prominent issue which is being addressed by countries all around the world through a wide-variety of policy and legal strategies and tools, including conventions and treaties, such as the Paris Agreement of 2016. Binding international agreements are created to ensure a combined effort to limit the effects of climate change on the world. These actions are furthered by individual countries, who either ratify these conventions, or legislate their own goals on what they think is achievable, and individualise these goals to what their country can achieve.

In Denmark, they have set rather ambitious goals compared to the rest of the world. However, these goals are currently being achieved and on track with actions such as expanding renewable energy usage, and by the phasing out of petrol and diesel cars. This is being achieved through the extensive government funding for this issue, and the Danish Council on Climate Change providing concrete recommendations and reports to the government on victories as well as key improvements. In Australia, the goals set are less ambitious, but perhaps more practical in achievement due to current circumstances. For instance, sectors such as mining, which have high greenhouse gas emissions, are essential to the Australian economy. Australia’s goals at present include focusing majorly on fuel reductions. However, these conservative goals are still not being achieved, with the rate of reductions required to be doubled in the 5 years to 2030 and tripled in the decade to 2035 in order to achieve the country’s stated goals. Both countries also

have a shared goal related to their agricultural industry, aiming to use new innovations to make farming more sustainable. Denmark and Australia are two examples of how countries are combatting climate change respectively, and also of their collective responsibilities.
In this paper, the author has attempted to address the national obligations of Denmark and Australia in relation to climate change, summarise their respective advantages and disadvantages in terms of action taken to limit the effects of climate change, and conclude with an analysis to determine global best practices.

Climate change is undoubtedly one of the most prevalent issues of our time, which showcases how different countries react under pressure to fast changing situations. It remains to be seen if the issue will continue gaining as much attention and traction in the coming years, with focus quickly shifting to evolving concerns such as Artificial Intelligence governance etc. Regardless, our combined response to the threat of climate change will shape the future of this planet and many many generations to come.

Works Cited

Australian Government. “Australia-Denmark Strategic Partnership Arrangement - Vilnius, 11 July 2023.” Australian Government Department of Foreign Affairs and Trade, 2023, https://www.dfat.gov.au/geo/denmark/australia-denmark-strategic-partnership-arrangement-vilnius-11-july-2023. Accessed 8 July 2026.

---. “Australian Carbon Credit Unit Scheme | Clean Energy Regulator.” Cer.Gov.Au, 2024, https://cer.gov.au/schemes/australian-carbon-credit-unit-scheme. Accessed 8 July 2026.

Bertelsen, Linda. “The Broad Majority in the Danish Parliament Allocates 5 Billion DKK from Green Budgetary Leeway.” DBDH - the District Energy Go-to-Partner, DBDH, 16 Apr. 2024, https://dbdh.org/the-broad-majority-in-the-danish-parliament-allocates-5-billion-dkk-from-green-budgetary-leeway/. Accessed 8 July 2026.

“Broad Majority in the Danish Parliament Distributes Approximately 5 Billion Kroner from the Green Room for Manoeuvre.” FM.dk, 15 Apr. 2024, https://fm.dk/nyheder/nyhedsarkiv/2024/april/bredt-flertal-i-folketinget-fordeler-ca-5-mia-kroner-fra-det-groenne-raaderum/. Accessed 8 July 2026.

Boya Jiang, Danting Fan, Dimitri de Boer, and Xiaoyu Zhang. Landmark Climate Change Cases. July 2022, https://www.clientearth.org/media/q0jak2fr/10-landmark-climate-change-cases_clientearth_compressed.pdf. Accessed 8 July 2026.

CCPI. “Climate Change Performance Index.” Climate Change Performance Index, 2025, https://ccpi.org/. Accessed 8 July 2026.

Climate Change Authority. “2025 Annual Progress Report | Climate Change Authority.” Climatechangeauthority.Gov.Au, 2025, https://www.climatechangeauthority.gov.au/2025-annual-progress-report. Accessed 8 July 2026.

---. “About the Authority | Climate Change Authority.” Climatechangeauthority.Gov.Au, 2018, https://www.climatechangeauthority.gov.au/about-cca. Accessed 8 July 2026.

Climate Change Performance Index. “Ranking | Climate Change Performance Index.” Climate Change Performance Index, 2025, https://ccpi.org/ranking/. Accessed 8 July 2026.

Climate Council. “What Does Net Zero Emissions Mean?” Climate Council, 14 Apr. 2023, https://www.climatecouncil.org.au/resources/what-does-net-zero-emissions-mean/. Accessed 8 July 2026.

DENMARK - National Targets and Contributions Foreseen in the Draft National Energy and Climate Plan. 2021, https://energy.ec.europa.eu/document/download/7d52df51-97ab-4562-97d9-692123ad0272_en. Accessed 8 July 2026.

“Denmark (DNK) Exports, Imports, and Trade Partners.” OEC World, Mar. 2026, https://oec.world/en/profile/country/dnk. Accessed 8 July 2026.

Directorate-General for Communication. “Denmark - Final Updated NECP 2021–2030 (Submitted in 2024).” European Commission, July 2024, https://commission.europa.eu/publications/denmark-final-updated-necp-2021-2030-submitted-2024_en. Accessed 8 July 2026.

Fluence, Tessa. “Legal Explainer: Environment Law Reform and the ICJ.” Environmental Justice Australia, 27 Oct. 2025, https://envirojustice.org.au/what-the-icj-ruling-means-for-australias-environment-laws/. Accessed 8 July 2026.

German Council on Foreign Relations. “Common but Differentiated Responsibilities (CBDR) | DGAP.” DGAP, https://dgap.org/en/research/glossary/climate-foreign-policy/common-differentiated-responsibilities-cbdr. Accessed 8 July 2026.

Glenn. “Australian Fossil Fuel Subsidies Growing Faster Than NDIS, Hitting $16.3 Billion in 2025–26.” The Australia Institute, 11 Mar. 2026, https://australiainstitute.org.au/post/australian-fossil-fuel-subsidies-growing-faster-than-ndis-hitting-16-3-billion-in-2025-26/. Accessed 8 July 2026.

“Gloucester Resources Limited v. Minister for Planning.” Climate Case Chart, Sabin Center, 2017, https://www.climatecasechart.com/document/gloucester-resources-limited-v-minister-for-planning_745d. Accessed 8 July 2026.

Justia. “Massachusetts v. EPA, 549 U.S. 497 (2007).” Justia Law, 2007, https://supreme.justia.com/cases/federal/us/549/497/. Accessed 8 July 2026.

Ministry of Foreign Affairs of Denmark. “Global Climate Action Strategy.” UM-ENEN, 2024, https://um.dk/en/foreign-policy/new-climate-action-strategy. Accessed 8 July 2026.

Morton, Adam. “Australian Governments Subsidising Fossil Fuel Use by More Than $30,000 a Minute, Analysis Finds.” The Guardian, 11 Mar. 2026, https://www.theguardian.com/environment/2026/mar/11/australian-governments-subsidising-fossil-fuel-use-by-more-than-30000-a-minute-analysis-finds. Accessed 8 July 2026.

NASA. “Evidence.” Science.Nasa.Gov, NASA, 23 Oct. 2024, https://science.nasa.gov/climate-change/evidence/. Accessed 8 July 2026.

Parliament of Australia. “Climate Change Bill 2022.” Parliament of Australia, 2022, https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6885. Accessed 8 July 2026.

Paulina Jabczyńska, Liselotte Jensen, and Ville Seppälä. “Roadmap to EU Climate Neutrality – Scrutiny of Member States.” Denmark’s Climate Action Strategy, 2024, https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/767173/EPRS_BRI%282024%29767173_EN.pdf.

Sustainable Development Solutions Network. Sustainable Development Report 2025: Financing Sustainable Development to 2030 and Mid-Century. Vol. 10, 2025, https://doi.org/10.25546/111909. Accessed 8 July 2026.

The Climate Council. “What Is Climate Litigation and How Can Australians Take Action?” Climate Council, 27 Apr. 2022, https://www.climatecouncil.org.au/what-is-climate-litigation-and-how-can-australians-take-action/. Accessed 8 July 2026.

Tso, Kathryn. “How Are Countries Held Accountable Under the Paris Agreement?” MIT Climate Portal, 8 Mar. 2021, https://climate.mit.edu/ask-mit/how-are-countries-held-accountable-under-paris-agreement. Accessed 8 July 2026.

UNCCD. “Convention.” UNCCD, https://www.unccd.int/convention/overview. Accessed 8 July 2026.

United Nations. “All about the NDCs.” United Nations, 2023, https://www.un.org/en/climatechange/all-about-ndcs. Accessed 8 July 2026.

---. “Convention on Biodiversity.” United Nations, 2021, https://www.un.org/en/observances/biological-diversity-day/convention. Accessed 8 July 2026.

---. “The Paris Agreement.” United Nations, 2015, https://www.un.org/en/climatechange/paris-agreement.

United Nations Environment Programme. “About Montreal Protocol.” UN Environment Programme, 2025, https://www.unep.org/ozonaction/who-we-are/about-montreal-protocol. Accessed 8 July 2026.

Structural Causes and Policy Responses to Teacher Psychological Distress in South Korea
By Taeyeon Kim

Abstract

There is a mental health crisis among South Korean teachers. Research has documented worrying rates of depression, burnout, anxiety, and stress among the nation's education professionals. This paper explores the nature of that psychological distress within the teaching profession, its causes, and what policy responses have been implemented to address the problem. Through an examination of a number of studies and surveys, the paper argues that the primary drivers of teacher psychological distress are structural rather than individual in nature. These negative pressures emerge from four areas: parental pressure, administrative overload, inadequate support for novice teachers, and rapid curriculum and digital changes. This means that while individual interventions (such as mindfulness training) are useful, the evidence suggests they are insufficient as a primary policy response. Korea's 2023 Teacher Rights Restoration Bills represent a meaningful legislative step in the right direction. However, administrative overload and novice teacher mentoring remain unaddressed by current policy. The paper concludes that sustainable improvement requires a twin response: consolidating existing legislative protections and expanding support in areas still found lacking.

Introduction

July 2023 marked a significant turning point in the way South Korean society perceived the teaching profession. That month, a 23-year-old elementary school teacher in Seoul took her own life due to intense work-related stress. Her death triggered unprecedented protests, with teachers taking to the streets demanding better protection and working conditions (Rashid). For the broader society, this was a shocking event that seemed to emerge from nowhere. However, for those within education, this death was the culmination of years of increasing pressure connected to higher workloads and other stressors. Peer-reviewed research supports this view. This research shows that South Korean teachers experience elevated rates of depression, burnout, anxiety, and stress due to a variety of factors. This paper examines what kind of policy responses are justified and what steps have been taken thus far. In assessing these questions, an analytical scaffold is constructed. Firstly, the issue of psychological distress among South Korean teachers is unpacked. Secondly, the key drivers of this distress are analyzed and discussed. Finally, a critical examination of policies and approaches that have been implemented is completed. The conclusion reached is that while individual interventions have some merit, the primary drivers of teacher psychological distress in South Korea are structural, and sustainable improvement requires policy responses that address those conditions directly. In some ways, policymakers have taken admirable steps to address these issues, but in other areas, more work is needed.

Psychological Distress among South Korean Teachers

Mental health difficulties faced by South Korean teachers have been documented across a variety of studies in recent years. Researchers have consistently identified elevated levels of

anxiety, depression, burnout, and stress within the profession. The scale and severity of the problem were foregrounded in Korean social consciousness when, in July 2023, an elementary school teacher in Seoul committed suicide due to work-related issues ("Teacher Suicide Exposes Parent Bullying in S Korea"). This incident was not surprising when examined in the context of research showing the declining mental health of education professionals over time. An examination of this research is instructive.

In 2024, a meta-analysis was published in SAGE Open that examined the findings of 201 related studies on burnout among Korean early childhood teachers (Xu et al.). A consistent theme throughout the studies examined was that many teachers were under psychological distress in their profession. This finding of a profession under considerable psychological strain is not surprising when put into the context of the many challenges facing the South Korean education system. Like many countries around the world, the COVID-19 pandemic was considered a disaster for the education system. In 2021, the journal *Frontiers in Psychiatry* published a survey of South Korean schoolteachers to explore how pandemic-related stress and anxiety were influencing educator mental health outcomes. Unsurprisingly, the study found a significant connection between COVID-related professional stressors and broader depression and subjective well-being among teachers (I. Cho et al.). While it could be argued that this was more a function of the pandemic, the study demonstrated how broader societal shocks reverberated within schools and how this had a distinct impact on teachers.

Other studies showed that it was not only during times of crisis that teachers' mental health seemed to suffer. In fact, more recent studies have shown that teacher stress levels did not simply return to the pre-pandemic baseline after the crisis passed. A 2024 study in OBM Neurobiology focused on middle and high school teachers during and after the COVID-19 pandemic. Within this study, teachers were screened for PTSD symptoms, anxiety, and depression. The conclusion of the study was stark: the stress teachers experienced during the time of crisis had a lingering impact on their psychological well-being. So, rather than the pandemic being a psychological aberration, it compounded pre-existing problems within the education space (J. Cho et al.).

Survey data also paints a similarly worrying trend for policymakers. In 2023, a survey was conducted by the Korean Teachers and Education Workers' Union, supported by Green Hospital in Seoul (Hoon). Nearly 3,500 teachers participated in the survey, which used the Center for Epidemiologic Studies Depression Scale as its measurement instrument. The results were alarming: 38.3% of educators showed "definite" depressive symptoms, a rate much higher than the estimated 8-10% of the general adult population. Further, 16% of respondents reported considering suicide at some stage. While it is important to acknowledge that this survey may suffer from response bias and the results were not vetted through peer-review, it is useful as a measure of showing the extent to which teachers themselves feel under immense strain. The results also align broadly with what the literature shows in terms of the mental health outcomes for South Korean educators.

When combined, these studies and surveys indicate that the teaching profession in South Korea is in crisis. There is genuine, measurable, and widespread psychological distress among teachers that is causing a number of negative symptoms. The most obvious question is then: what is creating this level of disharmony, and what can be done to address its causes?

Structural Causes of Teacher Psychological Distress

What the previous section makes clear is that many Korean teachers are under sustained psychological strain. This section unpacks the reasons for this distress occurring in the first place. A review of the literature on this issue points to the conclusion that the primary cause of this distress is not personal; it is structural. Four related causes are consistently mentioned as driving this problem: the collapse of teacher authority and unwarranted parental pressure, excessive administrative demands, poor support from those in authority for novice instructors, and pressures arising from constant policy changes.

A very prominent issue that is discussed in nearly all of the sources examined is the loss of teachers' authority in the classroom through the pressure of parents. Through threats of formal complaints or legal action, parents have inserted themselves into classes, leaving many teachers feeling powerless and threatened. Data drawn from the 2024 OECD Teaching and Learning International Survey (TALIS) is illustrative. Of the 3,477 Korean teachers who participated, 57% of respondents stated that addressing parental concerns and issues was a major source of work stress for them. This figure was substantially higher than the OECD average of 42% (OECD, Korea). In addition to this study, a 2022 survey from the Korean Teachers and Education Workers' Union found that 93% of teachers feared they may be accused of child abuse during their work. This is not an abstract fear, as of all teachers accused of this only 1.5% are eventually convicted (Gong). Many teachers see this system of complaints as a tool that has been weaponized by parents against them. The impact of having to work under the looming threat of legal action takes a real psychological toll, particularly when this is accompanied by verbal abuse and disrespect. A 2024 study published in *Behavioral Sciences* examined burnout among Korean secondary school teachers. In comparison to their international counterparts, Korean teachers were much more likely to report verbal abuse and conflict with student guardians as significant stressors ("How Do Korean Secondary School Teachers Perceive Psychological Burnout in Their Teaching Careers?"). The diminishment of self-esteem through this collapse of authority was seen as a strong driver of teacher burnout.

Excessive administrative work was also found to be a major structural driver of teacher unhappiness. The same TALIS data highlighted that half of Korean teachers stated that excessive administrative work is a significant stressor for them. Notably, Korean teachers were noted to spend more than six hours a week on administrative work, which is more than double the OECD average of less than three hours (OECD, Korea). Interestingly, this source of discomfort is not evenly distributed, and those teachers with more than ten years of experience were more likely to state that it is a work stressor. Crucially, the TALIS data also shows that among the variety of tasks that teachers need to complete, working on administrative tasks has the strongest association with lower well-being. In other words, the heavier the admin load, the unhappier the

teacher was. Administration work was seen as even more unfavorable than time spent grading or communicating with parents (OECD, "Results from TALIS 2024"). The administrative demands placed on Korean teachers are an important contributor to teacher unhappiness.

Rapid-fire changes to the school curricula and a focus on digital transformation have also been seen as core drivers of teacher psychological distress. In 2022, the Korean Ministry of Education made major curriculum changes and required implementation across primary and secondary schools from 2024 to 2027. This change forced teachers to substantially revise their teaching content and methods in response (Kim). In addition, Korea has committed $740 million to training teachers on so-called "AI-enhanced teaching practices" (NCEE). This shows that the education system is being remodeled and teachers are on the frontlines of these changes. Research shows that teachers in Korea usually resist these changes because they typically require large changes to existing teaching practices, which use up valuable preparation time. Teachers also complain that these changes are usually not well implemented and that training is minimal, meaning many educators feel unprepared for what the new curriculum requires (Kim). These pressures interact with other existing structural stressors; teachers who already feel stretched by the demands of their administration and students' parents now feel an added concern of being judged as incompetent when making the instructional changes required by the new curriculum's assessment framework or technology requirements.

Finally, there is the growing issue that many newly qualified teachers feel poorly supported while they find their footing in the profession. Only 12.5% of novice Korean teachers have some kind of assigned mentor to provide guidance and support as these new teachers begin their careers (OECD, "Results from TALIS 2024"). This places Korea 53rd out of 55 participating education systems and well below the OECD average of 26%. The majority of these new teachers are expected to learn on the job and cope with the challenges of the job as well as their more mature colleagues do. This is crucial because the early career period in education has the highest rate of burnout and attrition. TALIS data emphasizes that novice teachers who have an assigned mentor to whom they can reach out to have higher rates of job satisfaction. Without this support, these new teachers bear the full brunt of parental pressure, demands from the administration, and classroom management challenges, which can lead to burnout and psychological distress.

Individual Interventions: Evidence and Limits

Researchers and policymakers have explored a variety of focused interventions in an attempt to ameliorate the psychological distress issue. These programs are aimed at improving a teacher's coping capacity through mindfulness training, cognitive behavioral therapy (CBT), resilience programs, and more. This section analyzes what the evidence shows about the results of these interventions and evaluates whether they are sufficient as the chief policy response.

The idea of having focused individual interventions does have a research basis. For example, a 2023 review in the *International Journal of Environmental Research and Public Health* explored 40 studies conducted around the world and identified 16 psychological distress-related interventions that were used with teachers (Agyapong et al.). Of these 16

interventions, the most commonly studied were mindfulness-based interventions and Rational Emotive Behavioral Therapy (REBT). The results were positive, finding that such interventions did result in reductions in overall teacher stress. The breadth and scope of this review, which spanned education systems and countries, suggest that these programs do have merit.

Another, more rigorous examination was done by Beames *et al*. This 2023 systematic review and meta-analysis, published in *Educational Psychology Review,* identified 88 unique studies and focused on intervention programs that worked to lessen professional burnout among teachers. Results from randomized controlled trials were moderately encouraging: "In randomized controlled trials, the programs examined had large effects on stress (g=0.93), and moderate effects on anxiety (g=0.65), depression (g=0.51), professional burnout (g=0.57), and wellbeing (g=0.56)" (Beames et al.). These effect sizes show that a well-designed psychological support program can lead to improvements in teacher mental health outcomes.

That said, the study does also raise concerns. Firstly, the authors noted that the methodological quality of studies in the field was not of a high standard, particularly among non-randomized trials. There was also the fact that each study seemed to have its own unique approaches and strategies, making broad conclusions difficult. Finally, and probably most problematically, many of the interventions were judged to be too time-intensive to be feasible for working teachers. When placed into the context of the teachers discussed in the previous section — heavy administrative loads, parental pressure, curriculum change demands — it seems that these are the exact educators who would need this support most, but are the ones least likely to have the time to engage meaningfully with programs that require sustained additional time and effort outside of their working day.

Placing individual interventions at the forefront of policy response is also problematic in other ways. Much of the literature discusses tools to improve teacher resilience, which has been shown to reduce burnout and other signs of psychological distress. However, if the core stressors are structural (parental pressure, administrative overload, inadequate support, policy churn, etc.), then improving resilience will not remove those stressors, but only raise the bar in terms of what teachers are able to endure. Focusing only on individual coping does a disservice to teachers because it reframes psychological distress as a "personal deficit," rather than a very rational response to a toxic work environment. This does not mean individual support is unneeded. Rather, it is important to frame these individual support programs as complements to structural reforms and changes, not as substitutes for them.

Structural Policy Responses

What the previous sections have vividly illustrated is that the psychological distress among South Korean teachers is a real problem that has, at its core, a number of structural drivers. This means that individual interventions can be helpful to lessen some of the negative repercussions, but they cannot address the core structural issues. This section analyzes the existing policy responses to the crisis in teacher morale, then goes further by assessing what additional structural changes suggested by recent literature might be even more helpful.

Probably the most significant move by policymakers has been the Teacher Rights Restoration Bills (Rashid). The passage of these Bills followed weeks of protests by teachers after a series of teacher suicides. In response, the National Assembly made three important changes to relevant legislation. Firstly, teachers accused of abuse would no longer face mandatory suspension while investigations were conducted. For a suspension to occur, a minimum level of proof would be required. Secondly, financial support would be provided to teachers fighting lawsuits from parents. Finally, principals were given expanded powers to protect their staff members against frivolous claims by parents and unruly students. In addition, guidelines were published asserting the right of teachers to remove disruptive students from classes, and the Seoul Metropolitan Office of Education put in place a system to record all calls from parents to teachers. Procedures were also put in place to manage parental complaints before they reach teachers ("S Korea Passes New Law").

These reforms do seem to be a meaningful step when examined in the context of the structural issues identified earlier. In particular, the legislative and policy changes seem to take meaningful steps to address something that has been an acute stressor for teachers, according to the TALIS data and KTU survey, which is false child abuse allegations. In requiring a higher burden of proof for suspension and providing legal and financial support, policymakers have effectively balanced the rights of parents to be heard and those of teachers to work in a workplace free of frivolous harassment. What is also notable is the shifting of more responsibility for managing parental concerns onto principals as a way to protect staff. This seems to directly address previously reported concerns that there was "no institutionalized response system or obligation for school leadership's involvement" when parents made complaints (Gong). All these efforts reflect real structural interventions that take meaningful steps to ameliorate the issue.

That said, these steps leave two of the structural issues unaddressed. There have been no efforts to reduce administrative overload, which was identified by half of Korean teachers as a significant source of stress. Similarly, the plight of novice teachers needing mentorship and guidance is also absent from the legislative and policy changes. These represent significant failings that should be addressed in some way. If, as the data shows, administrative burden is the second-most commonly reported source of stress among Korean teachers *and* the issue most strongly associated with lower wellbeing, then a policy response that does not take steps to fix the issue can only be considered partial, at best.

The literature is instructive in terms of what steps could be taken to address these gaps. When it comes to administrative burden, there needs to be a workload redesign. As mentioned, TALIS data shows that Korean teachers are spending over six hours per week on administrative tasks, while simultaneously spending fewer hours on actual classroom teaching than the OECD average (OECD, Korea). This is the opposite of the ideal and needs to be addressed. Efforts need to be undertaken to streamline paperwork requirements, redistribute non-teaching tasks to support staff, and set reasonable limits on the expected administrative loads teachers are meant to complete. When it comes to novice teacher support, expanding mentoring provision has

substantial support in the literature (Mwesigwa and Nakato). TALIS data confirms that novice teachers with assigned mentors report higher job satisfaction and wellbeing (OECD, "Results from TALIS 2024").

Further, a broader point needs to be made on this issue. In the *SAGE Open* meta-analysis, one of the most powerful factors protecting teachers against burnout and stress is the sense of professional calling. This is closely linked to feeling respected and valued within a community. While the Teacher Rights Restoration Bills protect teachers after complaints are made, it does not change the culture around the profession in a way that removes the daily disrespect, excessive demands, and inadequate support that weaken the feeling of a professional calling. A wider conversation within South Korean society needs to occur around the teaching profession and the way teachers are treated and perceived. What this means is that current efforts need to go deeper into the problem. At the moment, the legislative and policy responses have gone some way in reducing teacher vulnerability by limiting some of the structural drivers of stress. However, a more comprehensive approach is needed to shift the way novice teachers are trained and the public's perception of the teaching profession as a whole.

Conclusion

South Korea's teachers are experiencing a genuine mental health crisis. The main drivers of this crisis are outside of their control, making teachers' jobs even more challenging. Parental pressure, administrative overload, the inadequate support of novice teachers, and constant policy change demands are all placing enormous pressure on teachers. While individual interventions do have a role to play in reducing these stressors, these are typically too time-intensive for already overloaded teachers. More importantly, building policy around individual remedies risks misdirecting policy from where the focus should be. Korea's 2023 Teacher Rights Restoration Bills are a good start because they directly address some aspects of the issues teachers face. However, administrative overload and lack of novice mentoring, both strongly evidenced as drivers of distress, remain unaddressed. While this review has attempted to be comprehensive, it does have its limitations. It draws exclusively on English-language sources and does not incorporate longitudinal evidence on the 2023 reforms, due to such studies not being completed yet. What current evidence does support is a twin response that sustains the legislative protections already enacted, while addressing the structural gaps that remain. Whether these steps are enough to restore the sense of professional calling that research identifies as the strongest protection against burnout is a more challenging question, but its level of importance signals the need for additional research.

Works Cited

Agyapong, Belinda, et al. "Interventions to Reduce Stress and Burnout among Teachers: A Scoping Review." International Journal of Environmental Research and Public Health, vol. 20, no. 9, Jan. 2023, p. 5625. www.mdpi.com, https://doi.org/10.3390/ijerph20095625.

Beames, Joanne R., et al. "Intervention Programs Targeting the Mental Health, Professional Burnout, and/or Wellbeing of School Teachers: Systematic Review and Meta-Analyses." Educational Psychology Review, vol. 35, no. 1, Mar. 2023, p. 26. Springer Link, https://doi.org/10.1007/s10648-023-09720-w.

Cho, Inn-Kyu, et al. "Schoolteachers' Resilience Does but Self-Efficacy Does Not Mediate the Influence of Stress and Anxiety Due to the COVID-19 Pandemic on Depression and Subjective Well-Being." Frontiers in Psychiatry, vol. 12, Oct. 2021. Frontiers, https://doi.org/10.3389/fpsyt.2021.756195.

Cho, Jihyun, et al. "Mental Health of Teachers during the First Wave of COVID-19 in Daegu, Korea: Challenges in Reopening Schools." OBM Neurobiology, vol. 8, no. 2, May 2024, pp. 1–29. www.lidsen.com, https://doi.org/10.21926/obm.neurobiol.2402221.

Gong, Se Eun. "South Korea Teachers Seek Protection from Harassment by Students' Parents." NPR, 5 Oct. 2023. Asia. NPR, https://www.npr.org/2023/10/05/1203824654/south-korea-teachers-seek-protection-from-harassment-by-students-parents.

Hoon, Moon Ki. "S. Korean Teachers at High Risk of Mental Health Issues, Survey Finds." The Korea Herald, 5 Sept. 2023, https://www.koreaherald.com/article/3206504.

How Do Korean Secondary School Teachers Perceive Psychological Burnout in Their Teaching Careers? https://www.mdpi.com/2076-328X/14/12/1210. Accessed 4 June 2026.

Kim, Jonghun. "Why Do Teachers Not Change While the National Curriculum Repeatedly Changes?: The 'Hidden' Resistance of Teachers in the Centralized System of Education in South Korea." International Journal of Educational Development, vol. 109, Sept. 2024, p. 103105. ScienceDirect, https://doi.org/10.1016/j.ijedudev.2024.103105.

Mwesigwa, Felix, and Nalukenge Nakato. "Mentorship in Teacher Education Programs: A Review of Practices, Outcomes, and Challenges." Acta Pedagogia Asiana, vol. 5, no. 1, 2026, pp. 15–29. tecnoscientifica.com, https://doi.org/10.53623/apga.v5i1.679.

NCEE. "Korea • NCEE." NCEE, 1 Jan. 2023, https://ncee.org/korea/.

OECD. Korea. OECD Publishing, 6 Oct. 2025, https://www.oecd.org/en/publications/results-from-talis-2024-country-notes_e127f9e2-en/korea_3d2c0051-en.html.

"Results from TALIS 2024: The State of Teaching." TALIS, Oct. 2025. www.oecd.org, https://doi.org/10.1787/90df6235-en.

Rashid, Raphael. "South Korea Passes Law to Protect Rights of Teachers after Mass Protests over Abuse from Parents." The Guardian, 22 Sept. 2023. World News. The Guardian,

https://www.theguardian.com/world/2023/sep/22/south-korea-teachers-rights-laws-mass-protests-parents-abuse.

S Korea Passes New Law to Protect Teachers from Bullying Parents. 21 Sept. 2023. www.bbc.com, https://www.bbc.com/news/world-asia-66877389.

Teacher Suicide Exposes Parent Bullying in S Korea. 3 Sept. 2023. www.bbc.com, https://www.bbc.com/news/world-asia-66655572.

Xu, Xiuhong, et al. “The Variables Associated with Burnout Among Korean Early Childhood Teachers: A Meta-Analysis.” Sage Open, vol. 14, no. 3, July 2024, p. 21582440241266993. SAGE Journals, https://doi.org/10.1177/21582440241266993.

The Effect of Financial Inclusion on Social, Economic, Political, and Environmental Indicators By Ardashir Prasetyo

Abstract

The paper investigates the impact of financial inclusion on various social, economic, political, and environmental indicators using data from 144 countries. Employing 18 different regression models, the findings indicated financial inclusion positively contributes to social welfare, economic growth, political stability, and environmental development. These results highlight the importance of financial inclusion as a key policy instrument for promoting sustainable development and reducing poverty across different regions of the world.

Keywords Financial Inclusion, Socioeconomic Indicators

Introduction

Financial inclusion is defined as the proportion of the population aged 15 years and above who have access to an account at a financial institution or with a mobile money service provider.

This study is motivated by the increasing importance of financial inclusion as a policy instrument for achieving sustainable development. Despite considerable progress in recent years, a substantial proportion of the global population remains financially excluded, particularly in developing countries. Limited access to financial services restricts individuals' ability to save, invest, and manage risk until they participate in economic activities. Consequently, financial exclusion creates barriers to inclusive growth and long-term development.

This study aims to examine the effect of financial inclusion on a broad range of social, economic, political, and environmental indicators using cross-country data. The findings indicate that financial inclusion is positively associated with GDP per capita growth, labor force participation rate, life expectancy, school enrollment, healthcare expenditure, income share lowest 20%, services value added, government effectiveness, rule of law, political stability, control of corruption, tax revenue, government consumption, and sustainability. In contrast, financial inclusion is negatively associated with unemployment, gini coefficient, income share highest 20%, and climate change.

This paper contributes to existing literature in several important ways. First, the paper provides a comprehensive analysis by examining the effect of financial inclusion on a wide range of social, economic, political, and environmental indicators within a single empirical framework. Unlike many previous studies that focus on a limited number of variables or a single dimension of development, this study integrates multiple development outcomes. Second, the paper expands the empirical literature by using cross-sectional data from 144 countries and estimating multiple models to capture the relationship between financial inclusion and various development indicators. Furthermore, this study provides evidence using the most recent data available from the World Bank's World Development Indicators (WDI) database.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature on financial inclusion and its effect on social, economic, political, and environmental indicators. Section 3 describes the data and methodology used in the analysis. Section 4 presents the empirical findings. Finally, Section 5 concludes the paper.

Literature Review

Various methodologies have been employed to investigate the effect of financial inclusion on social, economic, political, and environmental indicators. These methodologies include IV Probit Regression (Koomson et al., 2020), Cross-sectional Regression (Kim, 2016), GMM (Khan et al., 2022; Adediyan, 2020; Seifelyazal et al., 2023), Panel Data Models (Afonso and Blanco-Arana, 2024; Oz-Yalaman, 2019), FMOLS (Adeola and Evans, 2024), 2SLS (Islam et al., 2026; Edoh et al., 2021), OLS Regression (Balasubramanian and Kuppusamy, 2021; Vo et al., 2019), GLS (Koomson and Afoakwah, 2023), ARDL (Manasseh et al., 2023), Probit Model (Alhassan et al., 2019), Instrumental Variable methods (Ajide, 2020; Negera et al., 2024), Granger Causality analysis (Ozili, 2022), Dynamic Panel Data (Van et al., 2019).

The development indicators affected by financial inclusion also vary across the literature. These include agriculture share (Van et al., 2019), inflation volatility (Vo et al., 2019), renewable electricity output (Ozili, 2022), climate resilience index (Negera et al., 2024), tax revenue (Oz-Yalaman, 2019), corruption (Ajide, 2020), political stability (Alhassan et al., 2019), government effectiveness (Manasseh et al., 2023), Gini Index (Seifelyazal et al., 2023), health expenditure (Edoh et al., 2021), school enrollment (Koomson and Afoakwah, 2023), life expectancy (Adediyan, 2020), mean years of schooling for females (Balasubramanian and Kuppusamy, 2021), unemployment (Islam et al., 2026), exchange rate (Adeola and Evans, 2024), accessibility (Afonso and Blanco-Arana, 2024), human development (Khan et al., 2022), GDP growth (Kim, 2016), household size (Koomson et al., 2020).

Previous studies have also focused on different countries and regions. These include Ghana (Koomson et al., 2020; Koomson and Afoakwah, 2023), OECD countries (Kim, 2016), G20 countries (Khan et al., 2022), 46 least developed countries (Afonso and Blanco-Arana, 2024), Nigeria (Adeola and Evans, 2024), global samples (Islam et al., 2026; Balasubramanian and Kuppusamy, 2021; Alhassan et al., 2019), 14 West African countries (Adediyan, 2020), 17 African countries (Edoh et al., 2021), MENA countries (Seifelyazal et al., 2023), the COMESA region (Manasseh et al., 2023), Africa (Ajide, 2020), Ethiopia (Negera et al., 2024), and emerging countries (Vo et al., 2019).

Several studies have highlighted the positive relationship between financial inclusion and economic growth. For example, Van et al (2019) found that expanding financial inclusion is an effective strategy for promoting economic growth. Similarly, Afonso & Blanco-Arana (2024) reported that financial inclusion positively influences economic growth in least developed countries. However, although financial inclusion generally has a statistically significant and positive effect on economic growth, Vo et al. (2019) argued that its effects may become negative beyond an optimal level.

Many studies also emphasise the role of financial inclusion in improving quality of life. For example, Koomson et al. (2020) found that financial inclusion reduces both current poverty levels and the risk of households falling into future poverty. In addition, Koomson and Afoakwah (2023) reported that improved financial inclusion among parents enhances children' s learning outcomes and reduces delayed school enrolment. Adediyan (2020) found that while financial inclusion improves life expectancy, certain economic interactions may negatively affect life expectancy.

Financial inclusion is also considered an important tool for reducing income equality. For example, Kim (2016) found that improving financial inclusion is more effective in reducing inequality compared to tax progressivity. Similarly, Seifelyazal et al. (2023) confirmed that higher levels of financial inclusion are associated with lower levels of Gini Index.

Recent research has also emphasised the political impacts of financial inclusion. For example, Oz-Yalaman (2019) identified a positive relationship between financial inclusion and tax revenue. Furthermore, Ajide (2020) showed that financial inclusion contributes to reducing corruption in Africa, while emphasising the importance of strong institutions.

Other studies have highlighted the importance of financial inclusion in sustainability and climate related issues. For example, Negera et al. (2023) found that financial inclusion improves climate resilience among rural households in Ethiopia by helping them better cope with climate-related challenges. Similarly, Ozili (2022) argued that financial inclusion positively contributes to sustainable development through improvements in renewable energy use, productivity, and education. These findings suggest that expanding access to financial services can support environmental sustainability and broader development goals.

A summary of existing literature on the effect of financial inclusion is presented below:

Table 1. An Extensive Review of Literature on Financial Inclusion

Authors	**Years**	**Key Variables**	**Methodology**	**Country/ Region**	**Findings**	**Time Period**
Koomson et al.	2020	Financial Inclusion Household Factors	IV Probit	Ghana	FI reduces poverty.	2016/17
Kim	2016	GDP, Inequality, Unemployment, Inflation	Regression	OECD & EU	FI lowers inequality and supports growth.	2004-2011
Khan et al.	2022	Human Development, Trade, Population	GMM	G20	FI improves long-term stability.	2004-2017
Afonso & Blanco-Arana	2024	Financial Development, Unemployment	Panel Data	46 LDCs	FI increases economic growth.	1990–2021

Adeola & Evans	2017	Financial Access, Trade, FX	FMOLS	Nigeria	FI promotes economic diversification	1981–2014
Islam et al.	2026	Unemployment, RND, Population	GMM, 2SLS, PSM	Global	Digital FI reduces unemployment	2004–2020
Balasubramanian & Kuppusamy	2021	Female Literacy, Schooling	OLS, IV	Global	Female labor participation improves FI.	2017-2018
Adediyan	2020	Life expectancy	GMM	West Africa	FI improves life expectancy.	2010-2018
Koomson & Afoakwah	2023	Education Indicators	GLS	Ghana	FI improves children's educational outcomes.	2016
Edoh et al.	2021	Health Expenditure, Insurance	2SLS	Africa	FI increases health spending.	2008 to 2017
Seifelyazal et al.	2023	Inequality,, Unemployment	GMM	MENA countries	FI reduces income inequality.	2004-2019
Manasseh et al.	2023	Government, Volatility	ARDL	COMESA	FI has a positive impact on economic growth.	1997-2018
Alhassan et al.	2019	Political Stability, Education	Probit model	Global	Political stability increases FI.	2011-2017
Ajide	2020	Corruption, Inflation	Instrumental variable	Africa	FI reduces corruption when institutions are strong.	2005-2016
Oz-Yalaman	2019	Tax Revenue, Governance	Panel Data	Global	FI increases tax revenue.	2011-2017
Negera et al.	2024	Climate Resilience Index, Household Factors	IV	Ethiopia	FI improves climate resilience.	2018-2019
Ozili	2022	Renewable Energy, CO2 Emissions	Granger Causality	Global	FI supports sustainable development.	2001-2020
Vo et al.	2019	Inflation, Economic Stability	OLS	Emerging countries	FI improves stability.	2008-2015
Van et al.	2019	Agriculture, GDP	Dynamic Panel	Global	FI promotes economic growth, especially in low-income countries.	1980-2013

Empirical Findings

Data

The appendix presents data on financial inclusion and various social, political, and economic indicators obtained from the World Bank World Development Indicators database. The variables include Financial Inclusion, measured as account ownership at a financial institution or with a mobile money service provider as a % of the population aged 15 and above (Financial Inclusion Total), Secondary School Enrollment, measured as gross enrollment percentage (School Enrollment), Life Expectancy (Life Expectancy), Current Health Expenditure as a percentage of GDP (Health Expenditure), Gini Index (Gini Index), Income Share held by the lowest 20% of the population (Income Share Lowest 20%), Income Share held by the highest 20% of the population (Income Share Highest 20%), GDP per capita growth as an annual percentage (GDP Per Capita Growth), Unemployment as a percentage of total labor force (Unemployment), Labor force participation rate as a percentage of the total population aged 15 and above (Labor Force Participation), Services value added as a percentage of GDP (Services Value Added), Government Effectiveness (Government Effectiveness), Rule of Law (Rule of Law), Political Stability and Absence of Violence/Terrorism (Political Stability), Control of Corruption (Control of Corruption), Tax revenue as a percentage of GDP (Tax Revenue), General government final consumption expenditure as a percentage of GDP (Government Expenditure), Carbon dioxide emissions (Climate Change), Forest area as a percentage of land area (Sustainability).

Descriptive Statistics

Table 2. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Financial Inclusion Total	144	70.206	24.603	5.831	100
GDP Per Capita Growth	258	4.844	5.783	-22.584	33.769
Unemployment	150	7.614	5.306	.14	34.153
Labor Force Participation	140	59.875	8.738	34.573	87.4
Life Expectancy	265	71.431	7.482	40.279	85.533
School Enrollment	194	87.449	24.697	23.154	143.361
Health Expenditure	240	7.103	3.174	2.146	21.508
Gini Index	79	35.173	6.507	24.1	55.1
Income Share Lowest 20%	79	7.182	1.685	1.9	10.2

Income Share Highest 20%	79	42.666	5.177	33.3	59.8
Services Value Added	247	56.5	12.952	9.685	94.148
Government Effectiveness	205	-.032	.996	-2.397	2.248
Rule of Law	205	-.028	.998	-2.301	2.013
Political Stability	205	-.035	.999	-2.727	1.878
Control of Corruption	205	-.024	1.004	-1.837	2.333
Tax Revenue	155	16.033	5.979	0	35.2
Government Expenditure	217	17.536	7.411	2.36	61.033
Climate Change	249	492.481	3249.163	-75.286	49754.545
Sustainability	262	31.565	22.806	0	94.6

Table 1 presents the descriptive statistics for financial inclusion and various social, economic, political, and environmental indicators. Financial Inclusion consists of 144 observations, with a mean of 70.206 and a standard deviation of 24.603, ranging from a minimum of 5.831 to a maximum of 100. GDP Per Capita Growth includes 258 observations, with an average value of 4.844 and a standard deviation of 5.783, varying between -22.584 and 33.769. Unemployment has 150 observations, with a mean of 7.614 and a standard deviation of 5.306, ranging from a minimum of .14 to a maximum of 34.153. Labor Force Participation contains 140 observations, with an average of 59.875 and a standard deviation of 8.738, ranging from 34.573 to 87.4.

Life Expectancy includes 265 observations, with a mean of 71.431 and a standard deviation of 7.482, ranging from 40.279 to 85.533. School Enrollment consists of 194 observations, with an average value of 87.449 and a standard deviation of 24.697, ranging from 23.154 to 143.361. Health Expenditure has 240 observations, with a mean of 7.103 and a standard deviation of 3.174, ranging from a minimum of 2.146 to a maximum of 21.508.

Regarding income distribution indicators, Gini Index includes 79 observations, with a mean of 35.173 and a standard deviation of 6.507, ranging from 24.1 to 55.1. Income Share Lowest 20% has a mean of 7.182 and a standard deviation of 1.685 ranging from a minimum of 1.9 to a maximum of 10.2. Income Share Highest 20% has an average value of 42.666 and a standard deviation of 5.177, varying from 33.3 to 59.8.

Services Value Added consists of 247 observations, with a mean of 56.5 and a standard deviation of 12.952 ranging from 9.685 to 94.148. Government Effectiveness includes 205 observations, with a mean of -.032 and a standard deviation of .996 ranging from a minimum of -2.397 to a maximum of 2.248. Rule of Law has 205 observations, with a mean of -.028 and a standard deviation of .998 ranging from a minimum of -2.301 to a maximum of 2.103. Political

Stability has 205 observations, with a mean of -.035 and a standard deviation of .999 varying from -2.727 to 1.878. Control of Corruption consists of 205 observations, with an average of -.024 and a standard deviation of 1.004 ranging from -1.837 to 2.333.

Tax Revenue consists of 155 observations, with a mean of 16.033 and a standard deviation of 5.979 ranging from 0 to 35.2. Government Expenditure includes 217 observations, with an average value of 17.536 and a standard deviation of 7.411 varying from 2.36 to 61.033. Climate Change, measured through carbon dioxide emissions, has 249 observations, with a mean of 492.481 and a standard deviation of 3249.163 ranging from -75.286 to 49754.545. Sustainability, measured by forest area as a percentage of land area, includes 262 observations, with a mean of 31.565 and a standard deviation of 22.806 ranging from a minimum of 0 to a maximum of 94.6.

Overall, the descriptive statistics show that there is a lot of variation across countries on financial inclusion and selected social, economic, political, and environmental indicators. For financial inclusion, the average value is relatively high, but the standard deviation is quite large, so the difference in access to financial services is large between countries. Similarly, economic variables such as GDP per capita growth, unemployment, labor force participation rate, vary substantially, reflecting differences in economic development and labor market structures. The average values of social indicators, such as school enrollment and life expectancy are relatively high, indicating a broad improvement in global educational and health conditions. However, the wide ranges imply that significant disparities still exist among countries. Measures of income distribution also show differences in inequality across the sampled countries. Governance indicators average near zero. This is expected as these are standardized indicators. The wide ranges show that institutional quality varies widely between countries. Climate change has the highest standard deviation, showing carbon emission heterogeneity.

Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Where Y_i represents the dependent variable for observation *i*, including School Enrollment, Life Expectancy, Health Expenditure, Gini Index, Income Share Lowest 20%, Income Share Highest 20%, GDP Per Capita Growth, Unemployment, Labor Force Participation, Services Value Added, Government Effectiveness, Rule of Law, Political Stability, Control of Corruption, Tax Revenue, Government Expenditure, Climate Change, Sustainability.

β_0 represents the constant term, indicating the expected value of the dependent variable when the independent variable equals zero. β_1 is the coefficient for the independent variable, Financial Inclusion Total, and captures the change in the dependent variable associated with a one unit increase in financial inclusion. The term ϵ_i represents the error component, capturing the difference between the observed and predicted value of the dependent variable from the model. Specifically, the study estimates the following 18 regression models to examine the relationship

between Financial Inclusion and the selected social, economic, political, and environmental indicators:

$$\text{GDP Per Capita Growth}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 1)}$$
$$\text{Unemployment}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 2)}$$
$$\text{Labor Force Participation}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 3)}$$
$$\text{Life Expectancy}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 4)}$$
$$\text{School Enrollment}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 5)}$$
$$\text{Health Expenditure}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 6)}$$
$$\text{Gini Index}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 7)}$$
$$\text{Income Share Lowest 20\%}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 8)}$$
$$\text{Income Share Highest 20\%}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 9)}$$
$$\text{Services Value Added}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 10)}$$
$$\text{Government Effectiveness}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 11)}$$
$$\text{Rule of Law}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 12)}$$
$$\text{Political Stability}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 13)}$$
$$\text{Control of Corruption} = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 14)}$$
$$\text{Tax Revenue}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 15)}$$
$$\text{Government Consumption}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 16)}$$
$$\text{Climate Change}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 17)}$$
$$\text{Sustainability}_i = \beta_0 + \beta_1 \text{Financial Inclusion}_i + \epsilon_i \text{ (Model 18)}$$

Findings

Figure 1. Scatter plot graphics between financial inclusion and GDP per Capita Growth, Unemployment, Labor Force Participation, Life Expectancy, School Enrollment, Health Expenditure, Gini Index, Income Share Lowest 20%, Income Share Highest 20%

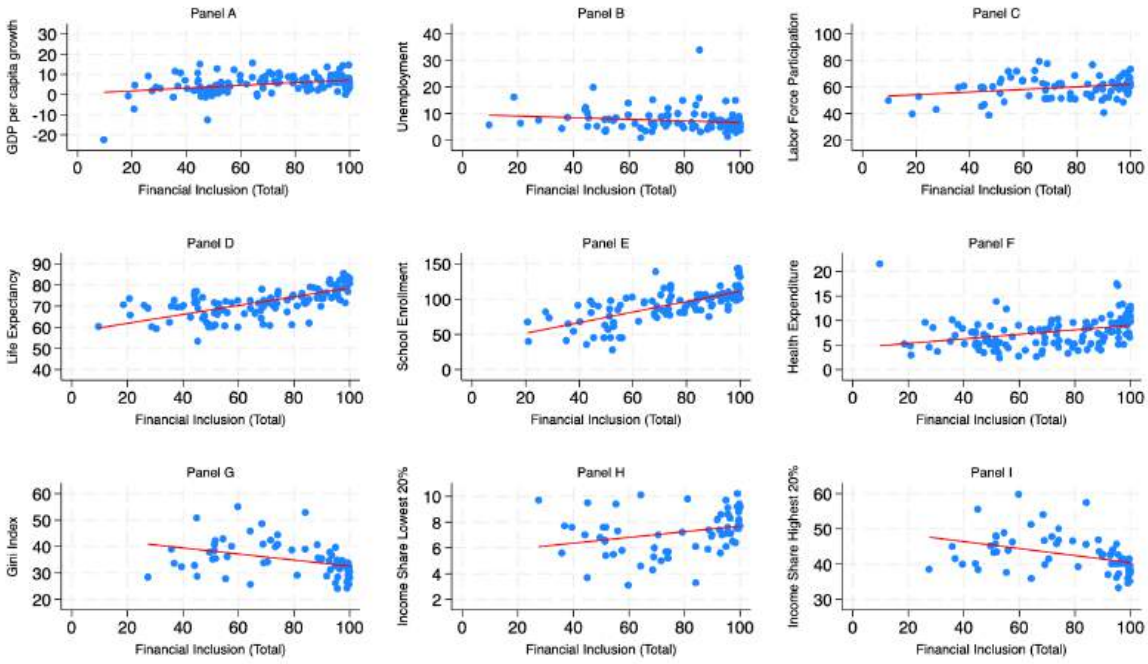


Figure 1 presents scatter plots illustrating the relationship between Financial Inclusion and 9 Social, Economic, and Political indicators. The findings indicate that Financial Inclusion is positively related to GDP Per Capita Growth, Labor Force Participation, Life Expectancy, School Enrollment, Health Expenditure, and Income Share Lowest 20%. This implies that higher levels of Financial Inclusion are associated with higher economic growth, greater labor force participation, better health and educational outcomes, and larger share of income going to the lower-income groups.

On the other hand, Unemployment, Gini Index, and Income Share Highest 20% are negatively related to Financial Inclusion. This means that the higher the level of Financial Inclusion, the lower the unemployment rate, the income inequality and the concentration of income among the wealthiest population group. Overall, the results provide evidence that financial inclusion can contribute positively to social and economic development outcomes, and reduce inequality related indicators.

Figure 2. Scatter plot graphics between Financial Inclusion and Services Value Added, Government Effectiveness, Rule of Law, Political Stability, Control of Corruption, Tax Revenue, Government Consumption, Climate Change, Sustainability

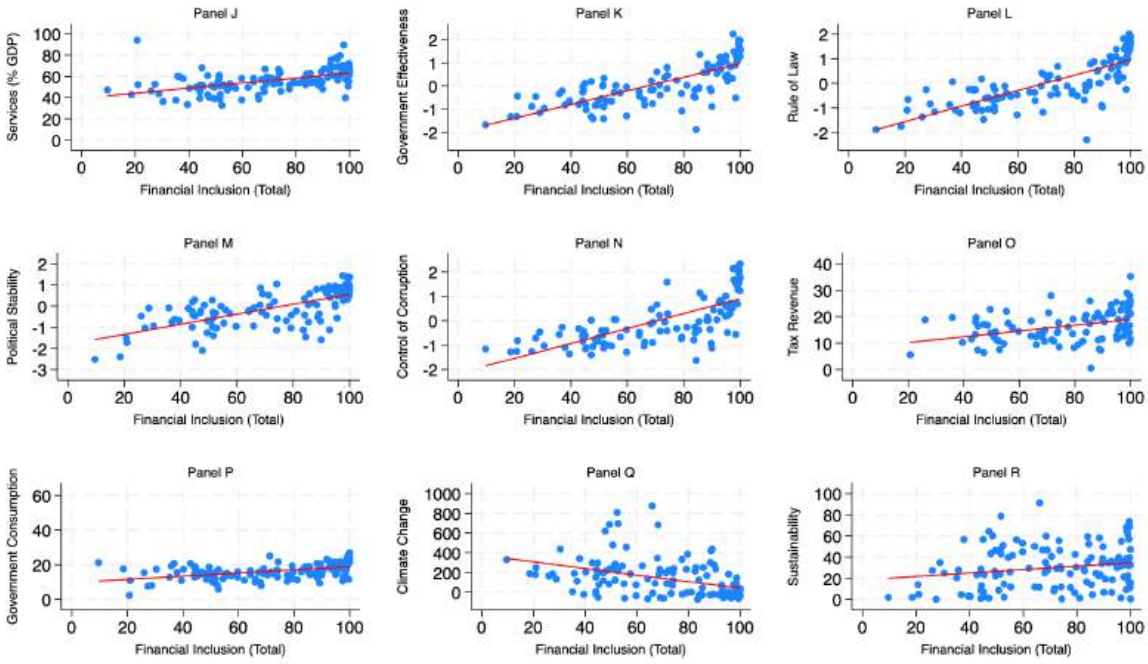


Figure 2 presents scatter plots illustrating the relationship between 9 Social, Political, and Economic indicators and Financial Inclusion. The results indicate that there is a positive correlation between higher levels of financial inclusion and Services Value Added, Government Effectiveness, Rule of Law, Political Stability, Control of Corruption, Tax Revenue, Government Consumption, and Sustainability. This shows that higher levels of Financial Inclusion contributes to higher levels of governance and higher sustainability. Financial inclusion is negatively correlated with Climate Change meaning that higher levels of Financial Inclusion are associated with lower Climate Change.

Table 3a. Regression Analysis

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
VARIABLES	GDP Per Capita Growth	Unemployment	Labor Force Participation	Life Expectancy	School Enrollment	Health Expenditure
Financial Inclusion Total	0.069*** (0.023)	-0.042* (0.023)	0.095** (0.040)	0.210*** (0.018)	0.757*** (0.072)	0.046*** (0.015)
Constant	0.368 (1.814)	10.773*** (1.991)	52.198*** (3.523)	57.623*** (1.459)	35.603*** (6.033)	4.352*** (1.129)
Obs	143	110	104	144	121	142
R-squared	0.125	0.038	0.065	0.510	0.515	0.129

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3a presents the result of six different regression models in which the dependent variables are GDP Per Capita Growth, Unemployment, Labor Force Participation, Life Expectancy, School Enrollment, and Health Expenditure. Each column represents a separate

regression model, while Financial Inclusion Total is used as the explanatory variable in all specifications. The paper uses t-values to determine the statistical significance of estimated parameters. We indicate the statistically significance of the result at one percentage level (***) five percentage level (**) and ten percentage level (*). The sign of estimated coefficient indicates the direction of relationship between explanatory and dependent variables, where "-" represents a negative effect, and "+" represents a positive impact on the dependent variable.

The results show positive relationships between Financial Inclusion Total and GDP Per Capita Growth, Labor Force Participation, Life Expectancy, School Enrollment, and Health Expenditure, while a negative relationship between Unemployment and Financial Inclusion Total. These findings suggest that greater financial inclusion is associated with improved economic and social outcomes, specifically countries with higher levels of financial inclusion tend to experience stronger economic growth, higher labor force participation, better health and education outcomes, and greater health expenditure. Furthermore, the negative relationship with unemployment indicates that greater access to financial services may contribute to reduced unemployment rates. Overall, the results provide evidence that financial inclusion plays a significant role in economic development and social welfare indicators.

Table 3b. Regression Analysis

	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
VARIABLES	Gini Index	Income Share Lowest 20%	Income Share Highest 20%	Services Value Added	Government Effectiveness	Rule of Law
Financial Inclusion Total	-0.111*** (0.034)	0.021** (0.009)	-0.096*** (0.026)	0.251*** (0.042)	0.030*** (0.002)	0.031*** (0.002)
Constant	43.835*** (2.973)	5.502*** (0.813)	50.133*** (2.298)	38.202*** (3.312)	-1.989*** (0.133)	-2.114*** (0.129)
Obs	71	71	71	142	122	122
R-squared	0.135	0.075	0.160	0.311	0.611	0.620

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3b presents the result of six different regression models in which the dependent variables are Gini Index, Income Share Lowest 20%, Income Share Highest 20%, Services Value Added, Government Effectiveness, and Rule of Law. The results indicate a positive relationship between Financial Inclusion Total and Income Share Lowest 20%, Services Value Added, Government Effectiveness, and Rule of Law. In contrast, a negative relationship is observed between Financial Inclusion Total and both Gini Index and Income Share Highest 20%. These findings suggest that greater Financial Inclusion is associated with improved income distribution and stronger institutional outcomes. Overall, the findings imply that financial inclusion plays an important role in promoting institutional quality.

Table 3c. Regression Analysis

	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
VARIABLES	Political Stability	Control of Corruption	Tax Revenue	Government Consumption	Climate Change	Sustainability
Financial Inclusion Total	0.024*** (0.002)	0.029*** (0.002)	0.106*** (0.023)	0.090*** (0.018)	-10.949** (4.248)	0.153** (0.064)
Constant	-1.834*** (0.184)	-2.070*** (0.139)	8.403*** (1.747)	9.916*** (1.428)	1,049.340*** (372.462)	19.496*** (4.907)
Obs	122	122	110	137	140	142
R-squared	0.468	0.545	0.151	0.216	0.086	0.037

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3c presents the result of six different regression models in which the dependent variables are Political Stability, Control of Corruption, Tax Revenue, Government Consumption, Climate Change, and Sustainability. The results indicate a positive relationship between Financial Inclusion Total and Political Stability, Control of Corruption, Tax Revenue, Government Consumption, and Sustainability. In contrast, a negative relationship is observed between Financial Inclusion Total and Climate Change. These findings suggest that higher levels of Financial Inclusion are associated with stronger institutional, fiscal, and environmental outcomes. Overall, the results imply that financial inclusion can contribute not only to economic and institutional development, but also to environmental sustainability.

Conclusion

This study examines the impact of financial inclusion on a broad range of social, economic, political, and environmental indicators, using a comprehensive empirical framework. Through regression analysis, this study provides empirical evidence on how financial inclusion contributes to overall societal development and progress.

The empirical findings suggest that financial inclusion plays a significant role in promoting development outcomes. Specifically, financial inclusion demonstrates a statistically significant and positive relationship with social welfare, economic growth, political stability, and environmental development. These results indicate that greater access to financial services is associated with improved educational and health outcomes, stronger economic performance, enhanced institutional quality, and more sustainable environmental conditions.

These findings provide important implications for policy makers, by suggesting that financial inclusion can serve as an effective strategy for achieving sustainable development.

Future studies may extend this research by employing panel data analysis to provide more robust empirical evidence over the time period. Additionally, future studies could divide the sample into developed and developing countries to examine whether the level of economic development influences the relationship between financial inclusion and development outcomes. Future studies could also explore the regional effect.

Works Cited

Afonso, A., & Blanco-Arana, M. C. (2024). Does financial inclusion enhance per capita income in the least developed countries?. International Economics, 177, 100479.

Adediyan, A. R. (2020). Is financial inclusion a determinant of life expectancy? Evidence from the West African Region. Studies in Business and Economics (SBE), 23(1), 22–37.

Adeola, O., & Evans, O. (2017). Financial inclusion, financial development, and economic diversification in Nigeria. The Journal of Developing Areas, 51(3), 1–15.

Ajide, F. M. (2020). Can financial inclusion reduce the presence of corruption? Evidence from selected countries in Africa. International Journal of Social Economics, 47(11), 1345–1362.

Alhassan, A., Li, L., Reddy, K., & Duppati, G. (2021). The relationship between political instability and financial inclusion: Evidence from the Middle East and North Africa. International Journal of Finance & Economics, 26(1), 353–374.

Balasubramanian, S. A., & Kuppusamy, T. (2021). Does female labour force participation contribute to better financial inclusion? Evidence from cross-country analysis. Journal of Economic and Administrative Sciences, 37(4), 643–658.

Edoh, O. M., Nchofoung, T. N., Anchi, O. E., & Munna, A. S. (2021). The impact of financial inclusion on household health expenditures in Africa (No. WP/21/080). AGDI working paper.

Hussain, A. B., Islam, M., Ahmed, K. J., Haq, S. M. A., & Islam, M. N. (2021). Financial inclusion, financial resilience, and climate change resilience. In the Handbook of climate change management (pp. 2085–2107). Springer.

Islam, F., Chowdhury, M. A. F., & Abdullah, M. (2026). Impact of digital financial inclusion and technological innovation on unemployment. Journal of Financial Economic Policy, 1–27.

Kim, J. H. (2016). A study on the effect of financial inclusion on the relationship between income inequality and economic growth. Emerging Markets Finance and Trade, 52(2), 498–512.

Khan, N., Zafar, M., Okunlola, A. F., Zoltan, Z., & Robert, M. (2022). Effects of financial inclusion on economic growth, poverty, sustainability, and financial efficiency: Evidence from the G20 countries. Sustainability, 14(19), 12688.

Koomson, I., & Afoakwah, C. (2023). Can financial inclusion improve children's learning outcomes and late school enrolment in a developing country?. Applied Economics, 55(3), 237–254.

Koomson, I., Villano, R. A., & Hadley, D. (2020). Effect of financial inclusion on poverty and vulnerability to poverty: Evidence using a multidimensional measure of financial inclusion. Social Indicators Research, 149(2), 613–639.

Manasseh, C. O., Okanya, O. C., Logan, C. S., Ede, K. E., Ejim, E. P., Ozor, S. N., & Okiche, E. L. (2023). Digital finance, financial inclusion and economic growth nexus in COMESA. Russian Law Journal, 11(5), 376–401.

Oz-Yalaman, G. (2019). Financial inclusion and tax revenue. Central Bank Review, 19(3), 107–113.

Ozili, P. K. (2022). Financial inclusion and sustainable development: an empirical association. Journal of Money and Business, 2(2), 186–198.

Seifelyazal, M., Salaheldin, A., & Assem, M. (2023). The impact of financial inclusion on income inequality. Open Journal of Social Sciences, 11(6), 255–274.

Van, L. T. H., Vo, A. T., Nguyen, N. T., & Vo, D. H. (2021). Financial inclusion and economic growth: An international evidence. Emerging Markets Finance and Trade, 57(1), 239–263.

Vo, A. T., Van, L. T. H., Vo, D. H., & McAleer, M. (2019). Financial inclusion and macroeconomic stability in emerging and frontier markets. Annals of Financial Economics, 14(2), 1950008.

The Role of the Right Triangle in Engineering, Construction, and Architecture: A Study of Theory, History, and Practical Application By Timurkhan Abdullaev

Abstract

In this study, an analysis of the significance of the right triangle in engineering, construction and architecture will be done. The analysis will not only encompass theoretical concepts such as the Pythagorean theorem but will delve into numerous practical applications. Through the combination of historical and contemporary engineering facts, it will be proved that right triangles have aided in the creation of structural stability and precise and accurate construction for centuries. Professional surveys and technical assessments have confirmed that, despite the availability of modern computer-aided designs, right triangle fundamentals remain an essential and widely used tool for ensuring building security and durability.

Introduction

Mathematical Foundations of Right Triangles

A right triangle refers to a type of a triangle that has an interior angle measuring 90 degrees. The side opposite to the right angle is referred to as hypotenuse while the remaining sides are referred to as the legs (Artmann, 1999). In case of an isosceles right triangle, the two legs have equal lengths hence the two acute angles have 45 degrees (Alperin, 2005). On the other hand, a scalene right triangle has unequal legs. A right triangle with integer side lengths is referred to as a Pythagorean triangle (Alperin, 2005). The sum of two acute angles should be equal to 90° while that of all three angles is equal to 180°.

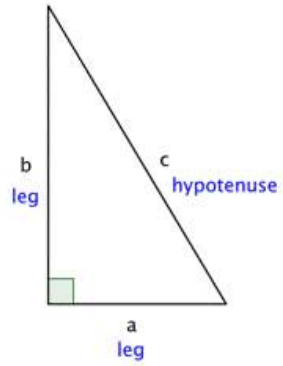


Figure 1. A right triangle, showing the two legs and the hypotenuse.

The main property in a right triangle is how the lengths of its sides relate to one another. The square of the hypotenuse is equal to the sum of squares of its legs. Mathematically expressed as $c^2 = a^2 + b^2$. This is the famous Pythagorean theorem. Converse of the theorem is if $c^2 = a^2 + b^2$ is true, then the triangle is right (Artmann, 1999).

Proofs of the Pythagorean Theorem

As is well known, there are many proofs of the Pythagorean theorem. Two geometric justifications are known. First is through the similarity of triangles. Drop the altitude from the right angle to the hypotenuse. This produces two smaller triangles, each similar to the larger one by two angles. From the equality of the ratios of the corresponding sides, the relationship $a^2 + b^2 = c^2$ follows (Shilets, 2024). The second method refers to the area decomposition method. Consider a square with side $a + b$, inside which four copies of the given right triangle are placed. The remaining central figure is a square with side $|a - b|$. The area of the larger square, $(a + b)^2$, is equal to the sum of the areas of the four triangles, $4 \cdot \frac{1}{2}ab$, and the area of the inner square, $(a - b)^2$. Equating these areas and cancelling like terms gives $c^2 = a^2 + b^2$. This construction is illustrated in the figure below: the square on the hypotenuse consists of the same areas as the squares on the legs plus four triangles (Shilets, 2024).

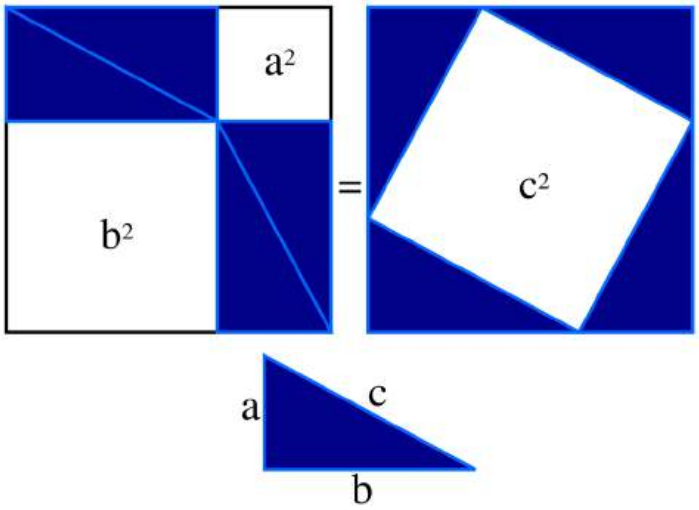


Figure 2. Geometric proof of the Pythagorean theorem through decomposition of figures.

Pythagorean theorem is used as a primary means of calculating the distance. Thus, while calculating the length of the diagonal of the frame in the rectangular distance with the horizontal distance between supports equal to 20 m and vertical distance equal to 15 m, we have $c = \sqrt{(20^2 + 15^2)} = 25$ m.

In the case of a right triangle there are three main trigonometric functions which are of great significance: sine, cosine and tangent. These functions help to calculate the lengths of the sides and angles of the triangle. Sine of an angle is the fraction of the hypotenuse which corresponds to the length of the side opposite to that angle. Cosine of an angle is the fraction of the hypotenuse which corresponds to the length of the side adjacent to that angle. Tangent of an angle is the ratio of the opposite and adjacent sides. In other words, sine and cosine compare the leg with the hypotenuse while tangent compares the legs to each other.

The Practical Significance of These Formulas

The theory of trigonometric functions and the Pythagorean theorem is not just about solving tasks in math classes but is rather a set of tools for addressing various practical problems. Trigonometry allows figuring out how long a staircase, diagonal, or roof will be and can be helpful in determining whether walls or a foundation make the necessary right angles, which is critical for a building to be level. Physics uses trigonometry to analyze movements and forces as when something is moving at an angle, its movement can be divided into two components - one horizontal and another vertical. Geography and navigation use the formulas to find the distance between two points and calculate the directions. These theories can be used to measure the height of any object knowing the distance to it and the angle of elevation. Simply put, trigonometry addresses slopes and angles, whereas the Pythagorean theorem deals with distances. All this information has extensive applications in life and different careers.

Historical Use of Right Triangles in Architecture

Throughout history, humans have been constructing their houses, temples, palaces, and fortresses striving for stability and level alignment. Angles had to be calculated precisely, and dimensions should be accurately measured. For this purpose, right triangles became a great instrument.

Ancient Egypt

Ancient Egyptians used detailed survey techniques for thousands of years in construction. They used ropes with 12 equally-spaced knots, which were spread flat on the ground and tied in knots 3, 7, and 12 (Alperin, 2005). Thus, a triangle with sides in the ratio 3:4:5 was formed and the angle between two smaller sides was perfectly right. Such an instrument could be used to create a perfect right angle without performing complicated mathematical calculations, which helped in laying out the pyramid's foundation and creating rectangular fields.

Ancient Greece and Ancient Rome

Both Greeks and Romans used geometry principles in the creation of harmonious architecture. For instance, the Parthenon in Athens was constructed with precision proportions and perfect symmetry (Yash, 2025). They carefully measured the columns' heights and distances to keep the façade even and proportionate. Also, the Romans continued this tradition. One may note the perfect mirror symmetry of the Roman Colosseum with even and symmetrical arches, galleries, and spectator's tiers (Carla, 2026).

Medieval Architecture

During the Middle Ages, magnificent Gothic cathedrals, which required accurate geometry. In order to accomplish this task, they used a system known as skeletal framework that consisted of large columns combined with pointed ribbed vaults and where lateral pressures were counterbalanced with the help of buttresses and flying buttresses (Early Byzantine Architecture. The Constantinopolitan School of Architecture | History of Medieval Architecture, n.d.). This

helped to make walls thinner and windows bigger without making them vulnerable to collapsing. For instance, one can speak about such buildings as Notre-Dame Cathedral that was built in Paris; it had many pointed arches and complicated stone vaults, each element of which was accurately combined in one whole. There were also other Gothic cathedrals that were built in a similar way; moreover, fortress walls and castles were also often constructed with right angles.

There is no doubt that throughout history, there were architects who tried to use the properties of right triangles and geometrical shapes in order to create a beautiful structure. For example, builders in Ancient Egypt, Greece, Rome, and during the Middle Ages used 3:4:5 ratio and symmetry (Artmann, 1999). Surveyors still use 3-4-5 triangles in order to lay out a right angle (Alperin, 2005).

Right Triangles in Structural Engineering

It is noteworthy that triangular structures withstand loads very well. When a load acts on the vertex of a triangle, it distributes through its sides down to the base without creating any bending of the components. Thus, the loads spread evenly among the structural components. For example, steel bridges have triangular structures; thanks to them, bridges can bear the load from heavy cars and trucks. The same applies to the roof of the house, whose rafters create triangles. Wind and snow loads act on the roof of the house, but thanks to triangles, the loads spread to the walls better.

In terms of rigidity, the triangle is the most rigid of the geometric figures because the only way to distort it is to change the length of its sides (Strong Structures With Triangles, 2017). In case when one of the sides of the triangle is subjected to external pressure, it experiences only compression and tension forces but no bending ones (Strong Structures With Triangles, 2017). Unlike the rectangle that can turn into a parallelogram under the effect of pressure, the triangle does not deform. This rigidity makes trusses very stable since materials resist tension and compression forces better than bending.

When the lengths of the sides of the triangle are equal to each other, then the triangle will not be deformed into another figure. Triangular structure will remain nonsagging. A truss consists of beams rigidly joined together and creating triangular cells (What Is a Truss? Common Types of Trusses in Structural Engineering, 2025). Triangles are present in all trusses in their different varieties.

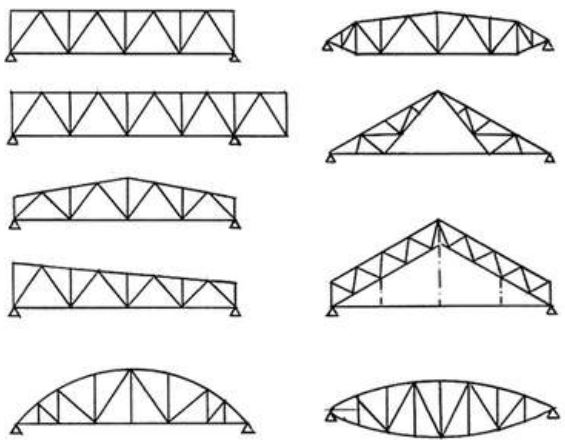


Figure 3. A truss composed of interconnected triangular cells.

Triangular trusses are widely used in bridge and roof construction. For example, steel bridges feature numerous triangular frames, which make the bridge lightweight yet strong. Rafter trusses in timber roof construction are also triangular, ensuring roof stability and rigidity. Triangular shapes are reliable in engineering. They distribute loads very effectively and maintain their rigidity. This explains the widespread use of triangular trusses and frames in roofs, bridges, towers, and other structures.

Rafter Systems and Roof Triangles

The rafter system refers to the supporting structure of a roof made of timber that transfers the loads generated by roofing materials, snow, and wind to building walls. There are several elements of the rafter system, namely, rafters (the sloping elements) and ties, braces, and struts connecting them. For example, the rafters in a gable roof meet at the ridge, resting on the wall plate (beam along the walls), which results in triangular formation. The design provides an opportunity for transferring the loads in a better way.

A gable roof design is one of the easiest roof designs. It features two sloping surfaces creating a triangular gable end. As literature shows, all variants may be described as triangles. Even a single-pitched roof may be considered as a right-angle triangle where one of its legs is the width of the building, the other leg – the height of the ridge, and the hypotenuse – the rafter. The triangular roof shape is practical and convenient since it provides better water and snow shedding and makes the process easier. In those areas where there is much precipitation and snowfall, a steeper triangular roof is used.

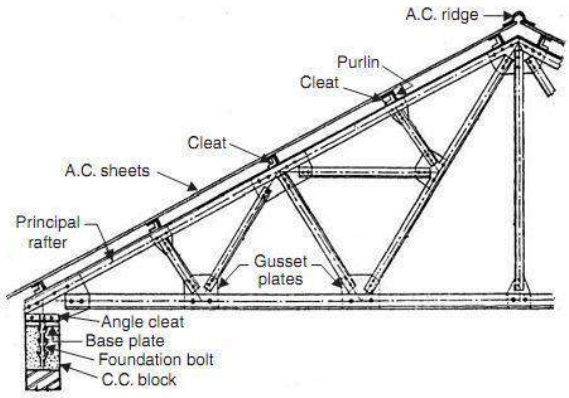


Figure 4. A gable roof, in which the rafters form a triangular frame.

It is possible to use a right triangle to calculate the height and length of rafters. As you know the width of the house and the slope of its roof, you will easily find the required data based on simple geometrical connections. No complicated equations are needed—suffices some understanding of triangles and angles. Triangle is a rigid figure that does not change its form when forces are applied because of the impossibility to distort it while its three sides remain unchanged. Such property of triangles is highly valuable for constructions. For example, the connection of two rafters in a suspended roof system forms a triangle that adds strength to the whole construction. Triangles are often used in construction systems, bridge construction and in support structures due to their constant resistance to the loads. Moreover, choice of the proper figure simplifies the process of installation. The right triangle with one 90° corner provides easy control of the plumb and level and the length of the slope can be calculated with the help of Pythagorean theorem as simply as the calculation. Hence, the rafter system can be assembled without much calculations due to the natural strength and right angle in the triangular framework. This is why the roofs of sheds and gazebos often have a triangular shape.

Design of Stairs and Ramps

When designing stairs, regulatory documents and ergonomic requirements are followed. Not only the dimensions and wall placement are considered, but, above all, safety and convenience. Therefore, designers select optimal step sizes and flight widths to minimize the risk of injury.

The slope angle is the most important parameter for a staircase. A comfortable slope is generally considered to be approximately 30–45°. Anything steeper than this makes the staircase uncomfortable and dangerous. For residential and public staircases, a gentler slope is chosen. This reduces the load on the user and makes the descent smoother.

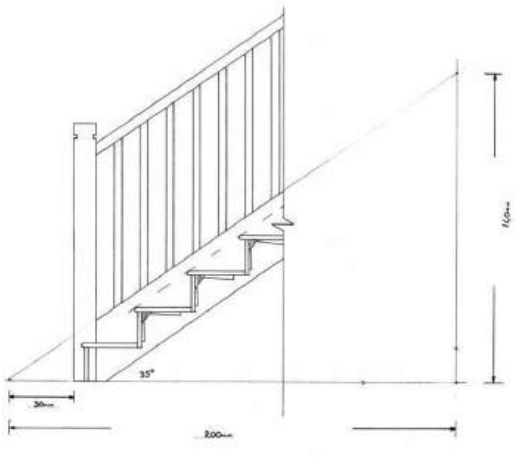


Figure 5. The slope of a staircase modeled as a right triangle.

First of all, it should be mentioned that ramps are made at a sufficiently low incline. The optimal degree is about 5° with a maximum of 9°. Due to a low incline, wheelchair-bound people can go up the ramp on their own, and the effort will not increase. In addition, it is necessary to place intermediate horizontal resting places: the greater the slope, the more frequently such rest places are needed.

As far as stairs are concerned, the main thing to understand is the geometry of the right triangle. So one side represents the total height, and another is the horizontal projection of the flight. Thus, the hypotenuse represents the length of the flight. On the basis of this data, it is possible to determine the necessary inclination and the general length of the object, thereby making it possible to build stairs adapted to the given space.

A certain number of requirements for safety and convenience appear. Thus, transverse slopes in steps and ramps are not allowed in the case of people with limited motor ability. Tread should be as horizontal as possible, and the longitudinal slope of the step should not exceed 10%. At the same time, surfaces of ramps and steps have to be slip-resistant. Ramps are usually made with the help of a fine-mesh lattice and coated by color contrasting with the surface. Stairs and ramps can be constructed considering such factors as the rise to be obtained, the available space, and relevant regulations. Engineers use the knowledge about the legs and hypotenuse of the right triangle in order to design flights.

The method of marking the right angle is called the Egyptian triangle. From the apex, it is necessary to measure 3 m on one side and 4 m on the other side and connect them with a rope. When the distance between the points equals 5 m, this means that the angle is right.

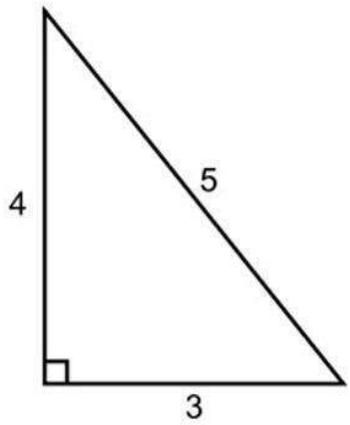


Figure 6. Marking a right angle on the ground using the 3-4-5 method.

To mark the angle, one can use a carpenter's setsquare, a tool with two rigid rulers joined at a right angle. One side is pressed against the already marked line, and a new line is drawn along the other edge of the square. The intersection of the edge of the square with the ground or a taut string is marked with chalk, or a stake is placed; this creates the second side at a right angle to the first. If a ready-made square is not available, a frame made of boards can be used instead: one side of the frame is placed along the first line, and a perpendicular is drawn along the other side of the frame. This creates a second marking line, perpendicular to the first.

In professional surveying, optical instruments are used to construct right angles. The most commonly used is the theodolite, an instrument with a telescope and angle markings. The theodolite is positioned over the angle and aligned, then the sight is aimed along one of the already marked lines. After this, the horizontal circle of the theodolite is rotated approximately 90° and the telescope is again aimed at the ground or target. A second point on the ground is marked using the sighting reticle, which is then used to draw a second marking line. This second line is perpendicular to the first. An optical level with a reticle can be used in a similar manner: it is first aimed at one line, then the sighting reticle is rotated and the intersection with the second line is marked. Both instruments provide highly accurate angle setting without the need for calculations.

A modern laser level allows getting a right angle fast with help of light beams. Numerous levels give two intersected beams at 90° creating a cross-line. The device is set up in such a way that one of the beams is parallel to the already marked side, and another automatically stands perpendicular to it. Being turned on, the lines on both sides are marked either with chalk, marker, or string put along the line. There are also rotary laser levels that give a visual right angle in case of the right-angle mode activation. The benefit of such a level is the automatic giving the right orientation and rapid marking of lines.

The stairs and ramps have a significant role in architectural design. They allow moving between different building levels and providing easy access to various places. In designing stairs and ramps, one has to take into account such parameters as safety, durability, and convenience.

The main elements of stairs design are several basic components among which there is the flight – the sloping part consisting of steps. The stairs also include landings, handrails, and elements fixing the system. All the steps in one flight have to be identical in size because the variety of step size may cause imbalance when a person walks the stairs. Moreover, the flow of people is also taken into consideration when designing stairs because in a building with many people, the stairs have to be wide enough to allow people to walk down simultaneously. Furthermore, in case of emergency, the stairs are the parts of escape routes, so they have to allow a quick and secure escape from the building. Ramps are used in cases when there is a need for accessibility for people with limited mobility, strollers and cargo transport. They are inclined planes with no steps allowing easy moving up and down between the levels of the building.

There is such an important parameter as slope in stairs and ramps' design, which influences the convenience and safety of moving up or down the stairs. The steep slope causes problems with walking and the high risks of falling while the gentle slope occupies a lot of space and is not effective in the building. For this reason, the building guidelines help in choosing the optimal slope. Stairs are designed with the aim to make people move in a natural way. The movement has to be rhythmic and easy.

Ramps usually have a more gentle slope than stairs in order to provide the safe moving of wheelchairs and people with limited mobility. The softer slope helps to reduce stress and improve the comfort, so the choice of slope plays an important role in the creation of a comfortable architectural environment.

In designing stairs and ramps, the engineers and architects use geometric models. One of the simplest models is the right triangle. Such a model allows understanding the structure of the stairs or ramp: the vertical part is represented by one leg, the horizontal base length by another leg, and the slope connecting the beginning and ending point is the hypotenuse. The use of the model allows understanding the slope, length of the flight, and the location of the steps. Even without using complicated calculations, one can see how the staircase will look and what place it will occupy in the building. Also, the diagram allows making comparisons of different variants and choosing the most convenient one.

Modern Architecture and Technology

Modern architecture has transformed due to technological development. In past times, everything was drawn manually on paper; today, almost all designing work is done on a computer. Such an innovation not only helps to accelerate the project creation process but allows to calculate dimensions with higher accuracy and predict the future appearance of the building (Brahambhatt, 2023). New materials and construction techniques have appeared which made it possible to build taller and innovative buildings. Modern technology makes it easier for architects and engineers to build a reliable and effective building (Ugah, Babalola, & Ekeh, 2024).

Nowadays, architects and engineers make use of special software programs for creating buildings. Technologies that are the most widespread among designers are CAD and BIM.

- CAD - Computer-Aided Design
- BIM - Building Information Modeling

These systems allow making precise computer drawings and thus creating designs very fast. It also becomes much easier to introduce changes into the plan since it is done on a computer. Moreover, BIM systems contain additional details that allow architects and engineers to see a digital model of a building rather than drawings. It means that all components of the building (walls, floors, pipes, electric installations etc.) are visible which gives the possibility to understand the functioning of the building after its completion.

Computer modeling allows us to test the design of the building prior to construction. Architects create a three-dimensional model to observe it from all sides and discover possible errors. Modeling allows testing the strength and behavior of the building depending on different circumstances such as distribution of loads and earthquakes. Thus, these possibilities increase the safety of buildings.

Modern construction uses lots of new materials that help to build lighter yet strong buildings. Composite materials, reinforced concrete and steel are used to build big bridges, complicated constructions and even skyscrapers. Furthermore, the modern design allows to make big interior spaces without using lots of columns. New technologies allow us to build environmentally friendly buildings.

Real-World Examples

Knowledge of theoretical information in architecture and construction is not enough – one should know about practical applications of theories. There are numerous landmarks, which demonstrate the application of geometry, triangles, and modern technologies. Some of them are presented below.

Bridges are considered to be very complicated engineering structures, since they should bear great weight and withstand wind pressure and other forces. Triangular shapes are often used in order to increase durability of bridges. Thus, trusses, the network of triangles connected with each other, are used in metal bridges quite often. Towers are designed as well as carefully calculated constructions, which should be able to withstand the wind pressure. To accomplish such a task, one should design the shape correctly and choose appropriate materials. Skyscrapers, which are extremely high buildings, are the same. As the higher building is, as stronger impact it receives from the wind. Engineers use special frames to protect skyscrapers from swaying and collapsing. Moreover, there are rigid frames inside and strong materials, such as steel and concrete, allow reaching great heights.

There are many world-famous landmarks, which demonstrate the use of the described information. Thus, the Eiffel Tower was built with the use of metal elements creating a strong frame and correct shape allowing bearing strong wind pressure. Another landmark, where the use of triangles and frames can be easily noticed, is large stadiums. The roofs of such stadiums are

designed with the use of frames, including triangles, in order to construct large space without additional supports.

Why Are Triangles So Effective?

Triangles are one of the most important geometric figures in building constructions and engineering. They may be found everywhere – from roofs and bridges to towers and stairs. There are several reasons for that, but the primary one is the fact that triangles cannot be distorted by any force. For example, a square might be easily deformed when subjected to lateral forces, turning into a parallelogram. On the contrary, a triangle remains the same. That is why engineers often use triangles for making structures more reliable.

Besides the stability, triangles help to distribute the loads effectively. If a certain force acts upon a building, the load will be distributed along the sides of the triangle to the foundation. It provides additional resistance to various external factors like heavy loads, strong winds, and even earthquakes. In addition to that, triangles are highly effective and useful when constructing roofs, bridges, and truss systems. Sometimes, it is possible to build such constructions using fewer materials.

Of course, triangles provide more than mere strength for construction. In the sphere of architecture, they can be used to create attractive and unique shapes of the buildings. Modern stadiums, towers, and glass houses often include triangles in their design.

Design Limitations and Errors

However, there are certain disadvantages associated with using triangles. For instance, the calculation errors or poor design may cause significant issues. One of the typical errors is when the wrong size or angles are selected to build the construction. Elements may be too thin or weak; hence, they will not bear the weight. Despite being durable and stable, the triangular construction will become unstable if it is made from poor quality materials. The other constraint is the complexity of some types of triangular forms. Unique architectural forms require special design and specific construction techniques; thus, the expenses of the construction can increase significantly. Moreover, engineers should take into consideration various natural conditions that can influence the building depending on its form and materials. Different shapes have different properties under different weather conditions. Hence, the same design that would work in one part of the world will not function in another area. Currently, the advancement of technology allows avoiding many errors. The computer models help engineers test the strength of construction and identify all the weak sides at the initial stage. Therefore, the triangle is one of the most reliable shapes to be used in architecture and engineering.

Results

Most participants use right triangles regularly in their work. In particular, 7 out of 10 respondents (70%) answered that they use right triangles “always,” “daily,” or “quite often.” This shows that right triangles are an important part of modern construction work. The majority of

respondents (6 out of 10, or 60%) confirmed that they use right triangles for general construction work. Moreover, two respondents indicated that they apply this approach in specialized work such as roofing and geodesy. This again confirms that right triangles are an essential part of construction work. Almost all respondents stated that geometry is mandatory (70%) or necessary (20%) for engineers, architects, and builders. Many of them gave clear answers such as “Of course,” “Required,” or “Critical.” Therefore, geometry is considered one of the fundamental disciplines for construction specialists. When asked about the projects in which right triangles are used, six respondents (60%) mentioned all projects, and two respondents (20%) highlighted their use in specific tasks. Therefore, right triangles are considered a universal tool, but also important for specific purposes. Most respondents believe that modern construction cannot be done without right triangles. The majority (60%) answered “No” when asked whether it is possible to avoid their use. Two respondents suggested that it might be possible in theory, but they were not fully sure. This means that right triangles are a very important part of practical construction and design. All the respondents agree that experience influences the preciseness of calculations, and one of them even said that "experience directly affects accuracy." Thus, experience enables people to make more accurate calculations. The skills in this sphere develop through education (3 of 10 respondents) and practice (8 of 10 respondents). One should note that people, who speak both about education and practice, have managerial and engineering positions, whereas those, who only speak about practice, are workers or builders.

Right triangles are widely used when making layouts. Seven respondents (70%) use triangles when marking and measuring the site of construction works; meanwhile, three respondents do not use triangles, since they are not engaged in practical construction work. When answering the question about the methods of checking the correctness of the angle of the triangle, five respondents (50%) named levels, two of them – squares and two – modern technologies such as lasers. One of the respondents mentioned the method "3-4-5". Interestingly, another respondent said that he/she checked "by eye." In general, the choice of the method depends on the specialization of the specialist. The most important professional qualities are knowledge and experience. People mention such skills as mathematics, physics, drawing, attentiveness, preciseness and practical experience.

The development of technology significantly changed the construction industry. About 40% of the respondents think that the construction process was better in the past; 20% believe that there are advantages and disadvantages; 20% say that construction is better now. Someone recalls that "walls were stronger before," and someone thinks that fastness of modern construction decreases its quality. To sum up, there is no unified opinion about this issue.

Concerning the last question about changes, most respondents (60%) said that these are technologies. There were four mentions of materials and three mentions that construction became faster. That means that, along with technologies and materials, construction became faster.

Conclusion

This research increased the knowledge about the role of right triangles in the field of engineering, architecture, and construction from different angles. Both the theoretical aspects of this figure and practical application have been analyzed. The review of literature and survey of experts' opinions demonstrate that despite the fast development of technology, this geometric figure is still a very important tool for designing and construction processes. Such theorems as the Pythagorean theorem and trigonometric functions allow calculating the height, distance, angle and other dimensions of the construction. It should be emphasized that these theorems allow solving practical problems and finding the applications in such spheres of activity as physics, construction, navigation and others that play an essential role in everyday life.

Analyzing history, we can say that right triangles are used for centuries. The ancient Egyptians, as well as the architects from Ancient Greece, Ancient Rome and the Middle Ages, used this figure to guarantee the stability and symmetry of buildings. From the interviews of the professionals, it becomes clear that there are a lot of geometrical methods that have remained relevant during thousands of years: the surveyors use "3-4-5 triangles" to make the exact right angle. One more major finding of the research is that triangular structure plays a crucial role in guaranteeing stability. The research proves that the triangle is the most rigid geometric figure and triangular trusses and frames are widely used in roofs, bridges, towers and other engineering constructions. In addition, the chapter about rafter systems emphasizes that triangular geometry guarantees the immediate stability and right angle which explains the popularity of the triangular solutions in the modern construction.

The survey of the professionals from the sphere of construction industry adds the special value for the research since it confirms its theoretical statements. According to the data obtained during the survey, seven out of ten respondents (70%) use right triangles "always," "daily" or "quite often". It means that right triangles play an essential role in the work of these specialists. In addition, the majority of respondents consider that geometry is a mandatory subject for engineers, architects and construction specialists. Besides, according to the survey results, the opinions of experts concerning the current development of the industry differ: some respondents consider that the current technologies increase the efficiency of the construction, but about 40% of experts consider that the quality of constructions was better in the past. The experts note that the current fast rhythm may influence the quality of the final construction. Thus, the objective of the research is fully accomplished. It proves that the right triangle is the fundamental element of the architecture, engineering and construction industry since the properties of this triangle are applied for structural analysis, marking, construction designs and creating engineering systems. Despite the fast development of technologies, the principles of geometry remain unchanged. Therefore, the right triangle is the essential element for developing reliable and safe construction and will continue playing this role in the future development of the construction industry.

Works Cited

Alperin, R. C. (2005). The modular Tree of Pythagoras. American Mathematical Monthly, 112(9), 807. https://doi.org/10.2307/30037602

Artmann, B. (1999). Euclid—The creation of mathematics. https://doi.org/10.1007/978-1-4612-1412-0

Brahambhatt, R. (2023, May 8). How modern architecture transformed our landscape, according to experts. ZME Science. Retrieved from https://www.zmescience.com

Carla. (2026, January 23). Symmetry in Roman and Greek architecture: design and precision. Retrieved from https://www.interiorsbycarla.com/symmetry-in-roman-and-greek-architecture/

Early Byzantine Architecture. The Constantinopolitan School of Architecture | History of Medieval Architecture. (n.d.). Retrieved July 7, 2026, from https://middleages.totalarch.com/universal_history_of_architecture/byzantium/early/constantinople

Shilets, A. (2024, July 9). Proof of the Pythagorean theorem, the formula, and real-life applications of the theorem. Retrieved from https://blog.tutoronline.ru/teorema-pifagora

Strong Structures with Triangles. (2017, June 13). [Video file]. Retrieved from https://www.pbs.org/video/strong-structures-triangles-qqlgej/

Surveyor's Mathematical primer. (n.d.). Retrieved July 7, 2026, from https://labs.landsurveyorsunited.com/math/

Ugah, U. U. K., Babalola, O., & Ekeh, E. (2024). The Shift from Traditional to Modern Architecture: A Review of 20th Century Development. Preprints.org. https://doi.org/10.20944/preprints202408.1679.v1

What is a Truss? Common Types of Trusses in Structural Engineering. (2025, October 20). Retrieved from https://skyciv.com/docs/tutorials/truss-tutorials/types-of-truss-structures/

Yash. (2025, May 23). The elegance of classical columns at the Acropolis. Retrieved from https://www.skurman.com/post/the-elegance-of-classical-columns-at-the-acropolis

The Correlation Between Environmental Conditions, Socioeconomic Vulnerability, and Health Outcomes in the Delridge Community, West Seattle, WA By Lucy Powell

Abstract

Key words: Environmental Justice, Water Quality, Water Quality Index, Socioeconomic Vulnerability, Health Disparities, Delridge, Bellevue

Environmental justice research evaluates how environmental burdens are commonly distributed disproportionately across vulnerable communities and the negative impact that it has on them. This study investigated the relationship between environmental conditions, socioeconomic vulnerability, and health outcomes by comparing two ZIP codes in the Seattle area: 98106 in the Delridge neighborhood of West Seattle and 98004 in Bellevue. Environmental data was acquired from King County's Water Quality Index (WQI) monitoring database, while the socioeconomic and health data was obtained from the Washington Tracking Network. The variables analyzed were stream water quality, housing affordability burden, and cancer mortality rates.

The results showed substantial differences between the two areas. Longfellow Creek in Delridge received a WQI score of 56, which is moderate environmental concern, while Yarrow Creek in Bellevue received a score of 85, meaning higher water quality and lower concern. Delridge also showed higher housing burden rankings (8-10 out of 10) and cancer mortality rankings (6-9) compared to Bellevue, which ranked substantially lower in both categories.

Although this study did not establish a direct correlation between the variables, the findings did suggest that environmental burdens, socioeconomic disadvantages, and negative health outcomes may overlap within vulnerable populations. The results highlight the importance of considering environmental quality, economic conditions, and public health all together when addressing the well-being of communities and environmental policy within them. Future research should expand the extent to which the impact the environment plays on communities.

Introduction

Longfellow Creek is a 4.2-mile urban stream that runs through West Seattle, specifically the Delridge neighborhood, and into the Duwamish Waterway. Delridge, the area through which the creek flows, has a long history of industrial development, beginning in the early 1900s with the Seattle Steel Company and later including railroads, mills, port development, and urban growth. In order for this industrialized neighborhood to be put in place, low flat coastal wetlands regularly flooded by tides, called tide flats, were filled, wetlands were drained, and much of the creek was diverted into underground pipes that altered the natural flow of the creek and eventually increased its vulnerability to pollution as well as the area around it (Bentley). Today, according to the King County Department of Natural Resources and Parks assessment on Longfellow Creek, about 93% of the creek's watershed is developed with roads, housing, shopping centers, and stormwater systems dominating land use in the Delridge community. This urbanization has led to runoff, sewer overflows, and legacy industrial contamination, all of which

are major sources of nutrients, suspended sediments, bacteria, and trace metals entering the creek (Gunderson et al.).

Demographics in the Delridge area have historically been working-class neighborhoods that have been shaped by immigration, industrial labor, and eventually demographic change. By 1990, more than one-third of the population was Filipino, Korean, Samoan, and African American, with an increase in renters and low-income households (*Delridge Snapshot*). To this day, Delridge remains a diverse and partially marginalized community that experiences displacement, industrial exposure, and infrastructure stress. Typically, communities with fewer resources often face greater exposure to environmental hazards, such as degraded water. Unfortunately, Longfellow Creek is a prime example of modern environmental injustice.

The water pollutants in Longfellow Creek affect community health in various ways, because it is designated for extraordinary-contact recreation and has a marvelous salmon spawning habitat. ("Salmon Surveys"). Pollution in the creek has both human and ecological consequences. Some key health pathways include fecal coliform and E. Coli, which increase the risk of gastrointestinal illness and waterborne disease. Nutrients and turbidity reduce oxygen levels, harming aquatic life and increasing flood and exposure risk, and the presence of trace metals can cause long-term neurological, respiratory, and developmental effects upon interaction. When considering how these pollutants directly affect the community, the Delridge Health Impact Assessment states that the most pressing issue concerning polluted water is its ability to transmit disease (Gunderson et al.).

Past research indicates that Longfellow Creek and polluted waterways like it have experienced long-term water-quality impairment, with persistent nutrient enrichment, bacterial contamination, and documented biological impacts. King County's Long-Term Monitoring Program has been sampling indicators such as temperature, dissolved oxygen, pH, turbidity, nutrients, suspended solids, and fecal coliform monthly since 1979. The program found that the creek qualifies as "impaired water" based on the temperature, dissolved oxygen, and fecal coliform categories (*DNDA.Org*).

Long-term trends show increasing nutrient enrichment and persistent bacterial contamination throughout the waters. While extensive monitoring exists, there remains a need for localized, community-based data to examine current conditions and equity implications in Longfellow Creek. To explore how environmental burden may vary between communities, this study compares the Delridge ZIP code 98106, which surrounds Longfellow Creek, with the Bellevue ZIP code 98004 inhabiting Yarrow Creek. Bellevue represents a contrasting community to Delridge, with different socioeconomic conditions and environmental infrastructure, providing a useful comparison for examining potential disparities.

This study investigates whether the environmental burden associated with water quality conditions is unevenly distributed across ZIP codes and how these differences relate to socioeconomic vulnerability and health outcomes. More specifically, this study seeks to understand whether areas with high rates of unaffordable housing, such as Delridge, are more burdened by environmental exposures that lead to poor health outcomes.

Materials/Methods

Data Collection

For this analysis, environmental conditions, socioeconomic vulnerability, and health outcomes were compared across two ZIP codes in the greater Seattle region. ZIP code 98106, located in the Delridge neighborhood of West Seattle, was chosen to represent a historically working-class and more marginalized community. ZIP code 98004, located in Bellevue, Washington, was selected as a contrasting area model due to its significantly higher income and low socioeconomic vulnerability.

Both the socioeconomic and health data were obtained from the publicly available website Washington Tracking Network (WTN), which is maintained by the Washington State Department of Health. The WTN database provides health, socioeconomic, and demographic indicators at multiple geographic scales across Washington State, based on ZIP codes and census tracts. The data was extracted from corresponding socioeconomic and health variables within the greater Delridge and Bellevue communities. Variables included the (1) percentage of households experiencing unaffordable housing (defined as spending more than 30% of household income on housing) and (2) reported cancer death rates within each ZIP code ("Washington Tracking Network").

Environmental data was obtained from King County's Water Quality Index (WQI) monitoring database. The WQI is a score used to evaluate overall water quality conditions based on multiple criteria. These include temperature, dissolved oxygen, pH, fecal coliform bacteria, total nitrogen, total phosphorus, total suspended sediment, and turbidity. For each ZIP code, the annual WQI score is calculated from a representative stream within the area. The Delridge ZIP code (98106) data was collected from monitoring Longfellow Creek, while water quality data from the Bellevue ZIP code (98004) was collected from Yarrow Creek ("King County's Water Quality Index Monitoring Database").

These data sets were compiled to compare environmental conditions, socioeconomic vulnerability, and health effects across the two ZIP code populations. The use of integrated variables allows for the examination of patterns and correlations across environmental burden and community vulnerability indicators experienced by marginalized communities. This approach emphasizes the relationship between environmental exposure and social vulnerability and how they interact.

Statistical Analysis

This study focuses on comparing two ZIP code populations using descriptive statistics and indicator values, rather than complex modeling.

Summary tables were created to compare environmental water quality, socioeconomic vulnerability, and health outcomes between the two ZIP codes, 98106 and 98004, respectively. The indicators examined included (1) WQI score for each watershed, (2) households experiencing unaffordable housing, and (3) cancer mortality rates.

The data were evaluated to determine whether the Delridge ZIP code represented a more socioeconomically vulnerable community and if it exhibited greater environmental burden/poorer health outcomes.

Results

Between the two ZIP codes that were examined, the water quality conditions differed substantially. The Delridge area (ZIP code 98106), which represented Longfellow Creek, received a water Quality Index (WQI) score of 56 out of 100, which falls within the moderate concern category. The Bellevue ZIP code (98004) received a WQI score of 85, indicating high water quality with low environmental concern ("King County's Water Quality Index Monitoring Database").

The data on housing unaffordability and cancer mortality rates are measured on a scale of 1-10, 10 represents the highest level of housing burden, while 1 represents the lowest.

In the Delridge ZIP code, the ranking of households experiencing unaffordable housing (defined as spending more than 30% of income on housing costs) generally ranged from 8 to 10 on a scale of 1-10. In comparison, ZIP code 98004 had lower housing-burden rankings, which typically range from 2 to 4 on the 1-10 scale.

As shown in Figure 2, ZIP code 98106 displays several census tracts have moderate-to-high cancer mortality rankings. It falls within the upper ranking categories of approximately 6 to 9 on the 1-10 scale. Some tracts show inconclusive data (as indicated by diagonal lines in the figures), which can occur in public health datasets when counts are too small to report. In contrast, the Bellevue ZIP code generally exhibited lower cancer mortality rankings, as shown in Figure 2.

Discussion

Overall

Across the indicators examined, the Delridge ZIP code showed lower water quality scores, higher housing burden rankings, and higher cancer mortality rankings compared with the Bellevue ZIP code, which experienced the contrary: higher water quality scores, lower housing burden rankings, and lower cancer mortality rankings. These results demonstrate consistent differences in environmental conditions, socioeconomic vulnerability, and health outcomes between the two communities.

Water Quality Index

The results indicate that the stream conditions that are associated with the Bellevue watershed are substantially healthier than those observed in Longfellow Creek.

Socioeconomic Indicators

As shown in Figure 1, almost all the census tracts within the Delridge community display higher ranking categories, indicating relatively greater housing affordability pressure compared with surrounding areas. The spatial distribution shown in the Figure indicates that most census tracts in the Bellevue area fall within the lower-ranking categories, suggesting lower levels of housing cost burden in this community, especially compared to the Delridge neighborhood.

Health Outcomes

Cancer mortality rankings varied between the two ZIP codes. In the Bellevue ZIP code, most census tracts fall in the lower categories of the ranking scale, which suggests lower overall cancer mortality levels, especially relative to the Delridge ZIP code.

Analysis

The results from this study indicate clear differences in environmental burdens, socioeconomic vulnerabilities, and health outcomes between ZIP code 98106, Delridge, and ZIP code 98004, Bellevue. Delridge had a lower WQI score, higher rankings for housing burden, and higher rankings for cancer mortality compared to Bellevue. While this study does not establish a definitive correlation between the datasets, the patterns observed suggest that environmental burden, socioeconomic stress, and health outcomes may be related in this geographic region.

One explanation for the patterns shown is the historical development that has taken place in Delridge, around Longfellow Creek. As mentioned previously, the Delridge area experienced extensive industrialization that began in the early twentieth century, including steel production, railway infrastructure, port development, and more. To support these activities, wetlands and tidal flats were filled, and portions of Longfellow Creek were eventually diverted into underground pipes in the neighborhood. These changes to the environment around the creek have increased the amount of urban runoff entering it and have contributed to long-term water quality impairment. In contrast, the Bellevue ZIP code examined in this study had a much higher WQI score, indicating comparatively healthier stream conditions and lower environmental stress overall in the area. Unlike Delridge, Bellevue developed with fewer large-scale alterations and is less industrialized, which may explain its healthier environmental conditions and higher WQI score.

For socioeconomic patterns, the census tracts within Delridge consistently ranked between 8 and 10 on the housing burden scale, indicating relatively high levels of unaffordable housing, especially compared to surrounding areas. On the other hand, the census tracts within Bellevue generally ranked between 2 and 4, suggesting lower housing cost pressure in the region. This represents roughly a four- to five-fold ranking difference between the two communities. A higher housing burden can increase community vulnerability, as households that are struggling economically may have fewer resources to address environmental risks, relocate, or invest in protective infrastructure. As a result, these socioeconomic disadvantages can intensify and compound the impacts of environmental stressors on communities, even when those hazards originate outside of the household.

Health data from this study showed a similar spatial pattern between the two populations. The Delridge neighborhood exhibited medium-to-high cancer mortality rankings, while Bellevue tended to fall within the lower ranking categories. First, it is important to note that cancer mortality is influenced by numerous factors, including age distribution, lifestyle factors, occupational exposures, and access to healthcare. Therefore, this study cannot conclude that poorer water quality directly caused higher cancer mortality rates. However, the observed overlap between environmental burden, socioeconomic vulnerability, and higher health risk rankings in the Delridge area suggests that environmental conditions may be part of a broader pattern of community stressors. Environmental research frequently finds that disadvantaged communities are more likely to experience environmental challenges. For example, an article about environmental justice and population health inequalities by Camila Alvarez and Clare Evans states, “Poor and minority communities are often disproportionately exposed to numerous environmental health hazards” (Alvarez and Evans).

The results of this study reinforce the importance of viewing Longfellow Creek not only as an ecological issue but also as a community and infrastructure issue, as reflected across various communities like it. Previous monitoring programs have documented water-quality impairment in the creek; this analysis contributes to understanding how watershed environments may influence other social and health indicators at the neighborhood level.

Limitations

There are several limitations that should be acknowledged when interpreting the results from this study. First, the analysis compares only two ZIP codes, which limits the ability to scale judgments of the findings and suggest generalizations for the entire region. Second, the socioeconomic and health variables used in this study are based on ranking systems, rather than direct household-level measurements, which introduces potential for inaccuracy or data limitations. Similarly, the Water Quality Index represents overall stream conditions but not every possible environmental exposure experienced by residents. Finally, the health data used in this analysis are observational and do not establish direct cause-and-effect relationships between environmental conditions and cancer outcomes; they only reflect potential correlations. Due to these limitations, the findings should be interpreted as a suggestive association, rather than direct proof of causation for the areas involved.

Ongoing projects

There are ongoing restoration and stormwater management projects that address environmental challenges facing Longfellow Creek, as recognized by local government agencies. For example, Seattle Public Utilities, in partnership with the Seattle Department of Transportation, has begun implementing Natural Drainage Systems (NDS) in several locations that are within Longfellow Creek. These systems use deep-rooted vegetation and engineered soils that are placed along roadway shoulders to capture and eventually filter polluted stormwater before it ever enters the creek. In addition to improving water quality, the project addresses

common environmental problems by reducing drainage issues, increasing vegetation, improving pedestrian access, and creating habitats for wildlife (*Longfellow Creek Natural Drainage System (NDS) Project*).

Moving Forward

Projects such as the Longfellow Creek NDS project are a prime example of organizations taking important steps toward improving watershed health. However, the patterns observed in this study indicate that environmental and socioeconomic disparities still exist within communities, even at a basic level. Continued investment in watershed restoration, stormwater management, and community infrastructure remains necessary to reduce long-term environmental burdens and hopefully improve health outcomes in the Delridge area.

Conclusion

This study has examined whether environmental burdens in and around Seattle's watershed areas are unevenly distributed across communities with differing levels of socioeconomic vulnerability and how those environmental conditions potentially influence health indicators. The analysis found that ZIP code 98106 in the Delridge Community has lower water quality scores, higher housing-burden rankings, and higher cancer mortality rankings compared with ZIP code 98004 in Bellevue, WA. While the study cannot establish direct causal relationships, the results suggest that the environmental burden overlaps with socioeconomic disadvantage and health risks.

The findings highlight the importance of considering environmental quality alongside economic conditions when evaluating community health. In this case, Longfellow Creek has experienced decades of urban development and watershed alteration, which have contributed to lengthy water quality challenges in the area. When these environmental conditions occur in communities already facing economic stress, the impacts of environmental degradation may intensify.

While action is needed, current restoration efforts, such as the Longfellow Creek Natural Drainage System project, provide a good example of local agencies addressing these environmental challenges through green infrastructure and stormwater management improvements. These projects help reduce runoff pollution and improve watershed health over time while also benefiting the surrounding community.

In conclusion, this research emphasizes that environmental conditions, community vulnerability, and public health should be considered together when comparing communities. Understanding how environmental burdens are distributed across neighborhoods can guide restoration efforts and ultimately help advance environmental policies, creating healthier and more equitable communities in the Longfellow Creek Watershed and beyond. Future research could expand on findings by examining other neighborhoods and additional datasets to understand how environmental burdens and potential health problems have developed over time.

Further studies could also research how historical development and industrialization have shaped present-day inequalities.

Supplemental information

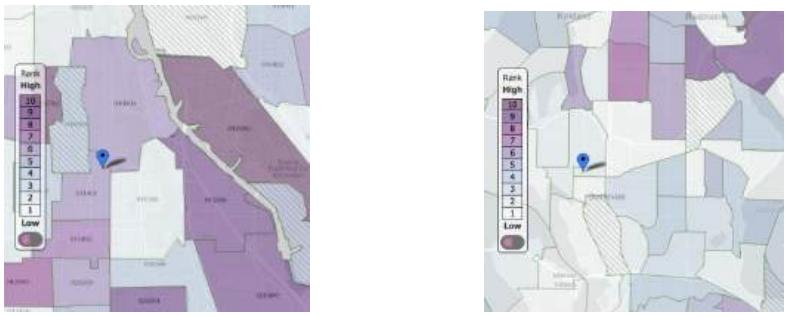
Zip Code	WQI Score	Unaffordable housing (Household spends > 30% of income)	Cancer Mortality Rates
98106	56 (Moderate)	8-10/10	6-9/10
98004	85 (High, low concern)	2-4/10	1-3/10

Table 1. Summary of environmental, socioeconomic, and health indicators for ZIP codes 98106 (Delridge) and 98004 (Bellevue).

Figure 1. Unaffordable housing burden rankings across census tracts in (L), ZIP codes 98106 in the Delridge neighborhood and (R), 98004 in Bellevue. Darker shading indicates higher housing burden rankings on a 1–10 scale. Census tracts in Delridge generally fall within the higher-ranking categories (approximately 8–10), indicating greater housing cost pressure, while Bellevue tracts mostly fall within lower categories (approximately 2–4), indicating comparatively lower housing burden.



Figure 2. Cancer Mortality Rankings across census tracts in (L), ZIP codes 98106 in the Delridge neighborhood and (R), 98004 in Bellevue. Darker shading indicates higher cancer mortality rankings on a 1–10 scale. Census tracts in the Delridge area generally fall within moderate to high-ranking categories (approximately 6–9), while census tracts in Bellevue fall primarily within lower ranking categories (approximately 1–3), suggesting lower overall cancer mortality rankings in the Bellevue region.



Works Cited

Alvarez, Camila, and Clare Evans. “Intersectional Environmental Justice and Population Health Inequalities.” Science Direct, 2021, https://www.sciencedirect.com/science/article/abs/pii/S0277953620307784.

Bentley, Judy. “Delridge History.” Log House Museum, https://loghousemuseum.org/history/delridge-history/. Accessed 4 Mar. 2026.

Delridge Snapshot. Aug. 2019, https://www.seattle.gov/documents/Departments/Neighborhoods/Districts/Neighborhood%20Snapshots/Delridge-Snapshot.pdf.

DNDA.Org. DNDA.org Loghousemuseum.org Seattle.gov/DelridgeHealthImpactAssessment Seattle.gov/DelridgeSnapshot Kingcounty.gov/watershedinfo. Accessed 3 Mar. 2026.

Gunderson, G., et al. Seattle’s Delridge Corridor Multimodal Improvement Project. 2015, https://www.seattle.gov/documents/Departments/OPCD/OngoingInitiatives/DelridgeActionPlan/DelridgeActionPlanHealthImpactAssessment.pdf.

“King County’s Water Quality Index Monitoring Database.” 2025, https://kingcounty.gov/en/dept/dnrp/nature-recreation/environment-ecology-conservation/science-services/streams-monitoring/water-quality-index.

Longfellow Creek Natural Drainage System (NDS) Project. 2025, https://www.seattle.gov/utilities/neighborhood-projects/longfellow-creek-natural-drainage-system.

“Salmon Surveys.” Puget Soundkeeper, 2025, https://pugetsoundkeeper.org/volunteer/salmon-surveys/.

“Washington Tracking Network.” 2025, https://fortress.wa.gov/doh/wtnibl/WTNIBL/.

Editing the Code of Life: CRISPR Technologies and the Future of Precision Medicine

By Esther Zhou[1]*

Abstract

CRISPR-Cas9 has transformed genome editing by providing an efficient, accurate and versatile method for modifying DNA. This review examines the history, molecular mechanisms, and therapeutic applications of CRISPR technologies while comparing traditional CRISPR-Cas9 with next-generation genome-editing approaches, including base editing and prime editing. It also explores the DNA repair pathways non-homologous end joining (NHEJ) and homology-directed repair (HDR), genome-editing strategies, and published studies demonstrating the clinical applications of CRISPR technologies. Traditional CRISPR-Cas9 remains an effective and versatile genome-editing platform for gene knock-out and knock-in applications through the NHEJ and HDR repair pathways. Base editing and prime editing improve editing precision by reducing or eliminating double-stranded DNA breaks, thereby minimizing unintended mutations. The published studies further demonstrate the therapeutic potential of CRISPR technologies in treating diseases such as Herpes Simplex Keratitis and Hereditary Tyrosinemia Type I. Although challenges including off-target effects, delivery efficiency, and ethical concerns remain, continued advances in CRISPR technologies are improving the safety and precision of genome editing. These developments position CRISPR as a promising platform for personalized medicine and future gene-based therapies.

Introduction

The emergence of CRISPR technologies redefines our ability to edit DNA with targeted precision (“CRISPR Made Simple”). CRISPR was originally discovered in bacteria as their natural defense mechanism against invading viruses, but has since been adapted as a powerful gene editing tool to help treat human genetic and non-genetic diseases (Gostimskaya). CRISPR technologies are used for numerous and various purposes, for example the most common use includes the use of the Cas9 protein which creates double stranded breaks in DNA to remove or insert edits (“CRISPR Methods”). Other technologies include editors which can precisely edit a single nucleotide, and prime editors which offer even greater flexibility and ability to edit DNA with more precision (“CRISPR Prime Editing Guide”).

This paper will delve into how CRISPR mechanistically works and how it's being used in medicine currently. We will compare different CRISPR tools, the DNA repair pathways they rely on such as non-homologous end joining (NHEJ) or homology-directed repair (HDR), and evaluate which tools are best suited for targeting specific diseases (“CRISPR 101: Non-Homologous End Joining”; “CRISPR 101: Homology-Directed Repair”). The clinical success of CRISPR depends not only on how well the editing tools work, but also on Non-Homologous End Joining”; “CRISPR 101: Homology-Directed Repair”). The

[1]Mayfield Senior School
*Corresponding author email: estherzhou1124@gmail.com

clinical success of CRISPR depends not only on how well the editing tools work, but also on how efficiently it can be delivered into cells within the body (Liu et al.). Ultimately, this paper will delve into the mechanisms behind CRISPR technologies and examine how they are transforming the treatment landscape for human disorders (Carroll).

History of Gene Editing

Early genetic engineering relied heavily on restriction enzymes, which are proteins that recognize short, specific DNA sequences and cut the DNA at or near those sites (“History of Genome Engineering”). Creating cuts in the DNA via restriction enzymes allowed scientists to edit the DNA sequence or even transfer genes between organisms (“History of Genome Engineering”). However, these enzymes displayed great limitations to scientists because they could only cut DNA at naturally occurring target sites, which meant scientists had little control over where a cut was made ("History of Genome Engineering"). This made the potential for editing DNA limited for precise gene editing, in addition to this technique already being fairly inefficient with high costs (“History of Genome Engineering”). Furthermore, scientists utilized naturally occurring Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) and further engineered the proteins to increase targeting range (“History of Genome Engineering”). These newly designed proteins could recognize and bind to a broader range of DNA sequences, allowing for more targeted cuts compared to restriction enzymes (“History of Genome Engineering”). ZFNs and TALENs were still relatively expensive to design for each new target, which limited the gene editing field (“History of Genome Engineering”).

In 2012, two scientists named Jennifer Doudna and Emmanuelle Charpentier adapted the CRISPR-Cas9 system into a powerful gene-editing tool ("CRISPR Timeline"). CRISPR is faster, easier, and cheaper than earlier methods because it does not require building new proteins for every target site (“CRISPR Made Simple”). Instead, researchers only need to design a short piece of guide RNA, which directs the Cas9 protein to any desired DNA sequence (“CRISPR Made Simple”). This reduced both the cost and time required for gene editing, making CRISPR accessible to laboratories worldwide (“CRISPR Made Simple”).

History and Mechanisms of CRISPR

CRISPR comes from a natural defense mechanism in bacteria (“CRISPR Made Simple”). When a virus infects a bacterium, the bacterium stores fragments of the viral DNA in its own genome in a special region called the CRISPR array . These stored fragments act like a genetic memory or a catalog of past infections (“CRISPR Made Simple”). If the virus invades again, the bacterium expresses the logged viral DNA from the CRISPR array and produces RNA molecules . These RNA molecules are complementary to the actively infecting viral DNA (“CRISPR Made Simple”). The RNA molecules bind to a protein called Cas9, and guide the Cas9 protein to the viral DNA via its complementary base pairing with the viral DNA. Cas9 then cuts the viral DNA, preventing further infection. Scientists adapted this natural system to cut DNA at precise locations in other organisms, turning CRISPR-Cas9 into a universal gene-editing tool ("CRISPR

Made Simple").

Traditional CRISPR-Cas9 editing works by creating a double-strand break in DNA, meaning that both strands of the DNA double helix are cut (“CRISPR Methods”). The advantage of a double-strand break is that it gives scientists the ability to disrupt genes or insert new DNA at precise locations, since the cell will attempt to repair the break using its natural repair systems (“CRISPR Methods”). In order to repair double-stranded breaks, the cell utilizes two main DNA repair pathways: non-homologous end joining (NHEJ) and homology directed repair (HDR) (“CRISPR 101: Non-Homologous End Joining”; “CRISPR 101: Homology-Directed Repair”) (Figure 1). However, this repair process has both pros and cons. If the cell uses NHEJ, the break is sealed quickly but often introduces small errors such as small insertions or deletions, which can turn off a gene but also create mutations in undesired locations, known as off-target edits (“CRISPR 101: Non-Homologous End Joining”). If the cell uses homology-directed repair, the break can be repaired more accurately with the help of a DNA template, allowing precise gene editing (“CRISPR 101: Homology-Directed Repair”). Due to the fact that double-strand breaks can sometimes cause harmful mutations, scientists developed two newer genome editing technologies known as prime editing and base editing. Both methods lower the mistakes caused by double-strand breaks, which is important for future medical treatments (“CRISPR Prime Editing Guide”).

Traditional CRISPR-Cas9 is still widely used because it is simple, inexpensive, and effective for either turning off genes or inserting new sequences with a repair template (“CRISPR Methods”). It remains the most versatile overall due to its ability to generate larger genetic changes and fewer size limitations (“CRISPR Methods”). However, the risk of errors from double-strand breaks makes it less ideal for treating human diseases (Wang et al.). Base editing is especially useful when a disorder is caused by a single-letter mutation in DNA, since it can precisely fix that error without creating a break (“CRISPR Base Editing Guide”). Prime editing is the most versatile of all, because it can not only correct single bases but also add or delete larger DNA segments, offering the potential to treat a wider range of genetic conditions (“CRISPR Prime Editing Guide”). Although prime editing is more technically complex, it represents an exciting step toward safer and more precise medical treatments (Wang et al.).

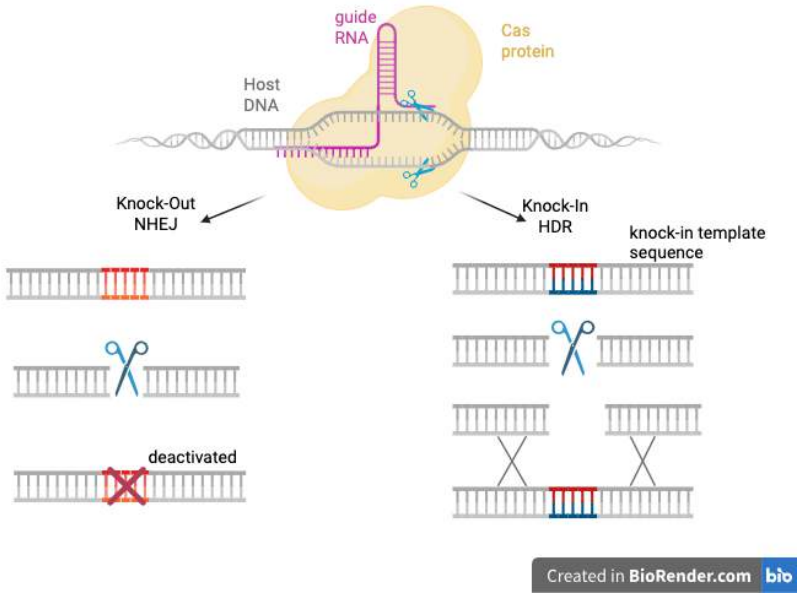


Figure 1. CRISPR-Mediated gene Knock-out and Knock-in.

The CRISPR-Cas system uses a guide RNA (gRNA) to direct the Cas protein to a specific target sequence in the host DNA, where it introduces a double-stranded DNA break. The break can be repaired by non-homologous end joining (NHEJ), which often generates insertions or deletions that disrupt gene function (gene knockout), or by homology-directed repair (HDR), which uses a donor DNA template to introduce precise genetic changes (gene knock-in). *Created in BioRender. Zhou, E. (2026) https://app.biorender.com/illustrations/6948cec47ab5054cfe921635*

Cas9's gene-editing power comes from its ability to create precise double-stranded breaks in DNA ("CRISPR Methods"). After the DNA is cut, the cell's repair systems determine the type of genetic modification that occurs ("CRISPR Methods").

CRISPR Editing Strategies: Knock-In and Knock-Out

CRISPR technology is commonly used for two main types of genetic edits: knock-out (KO) and knock-in (KI) ("CRISPR Methods"). In a knock-out, CRISPR-Cas9 cuts the DNA, and the cell repairs the break using non-homologous end joining, the cell's preferred repair pathway ("CRISPR 101: Non-Homologous End Joining"). NHEJ is error-prone and random, often causing insertions or deletions of one or more nucleotides ("CRISPR 101: Non-Homologous End Joining"). When these random mutations disrupt the gene's reading frame or introduce a premature stop codon, the gene becomes nonfunctional and is considered knocked out ("CRISPR Methods"). However, because NHEJ is uncontrolled, a random insertion or deletion can occasionally fix a mutation or restore gene function, which would result in a random knock-in rather than a knock-out ("CRISPR 101: Non-Homologous End Joining"). In contrast, guided

knock-in editing relies on homology-directed repair (HDR), in which scientists provide a DNA template that allows precise insertion or correction of genetic material ("CRISPR 101: Homology-Directed Repair"). The limitation with HDR is that it doesn't happen often due to how energy extensive it is for the cell, and usually only occurs in dividing cells ("CRISPR 101: Homology-Directed Repair"). While HDR gives scientists control over the exact edit, it occurs less frequently because cells favor NHEJ, requiring researchers to rely on chance or use chemical treatments to encourage HDR over NHEJ ("CRISPR Methods"). In a knock-out, CRISPR disables a gene by cutting the DNA and allowing NHEJ to introduce random mutations that prevent the gene from functioning ("CRISPR 101: Non-Homologous End Joining"). This approach is useful for studying gene function or eliminating harmful genes ("CRISPR Methods").

Base Editing vs. Prime Editing

Although traditional CRISPR-Cas9 is powerful, double-stranded DNA breaks can sometimes cause unintended mutations ("CRISPR Methods"). To address this issue, scientists developed base editing and prime editing ("CRISPR Base Editing Guide").

Base editing uses a modified Cas9 enzyme combined with a chemical enzyme to directly change one DNA base into another ("CRISPR Base Editing Guide"). There two main types of base editors: cytosine base editors (CBEs) and adenine base editors (ABEs). CBEs convert a cytosine (C) into a thymine (T) by chemically modifying the cytosine base, while ABEs convert an adenine (A) into a guanine (G) through a similar chemical reaction ("CRISPR Base Editing Guide") (Figure 2). Because CBEs and ABEs can only perform these specific base conversions, base editing is highly precise but limited in scope ("CRISPR Base Editing Guide"). This method is faster and safer for correcting single-letter mutations ("CRISPR Base Editing Guide").

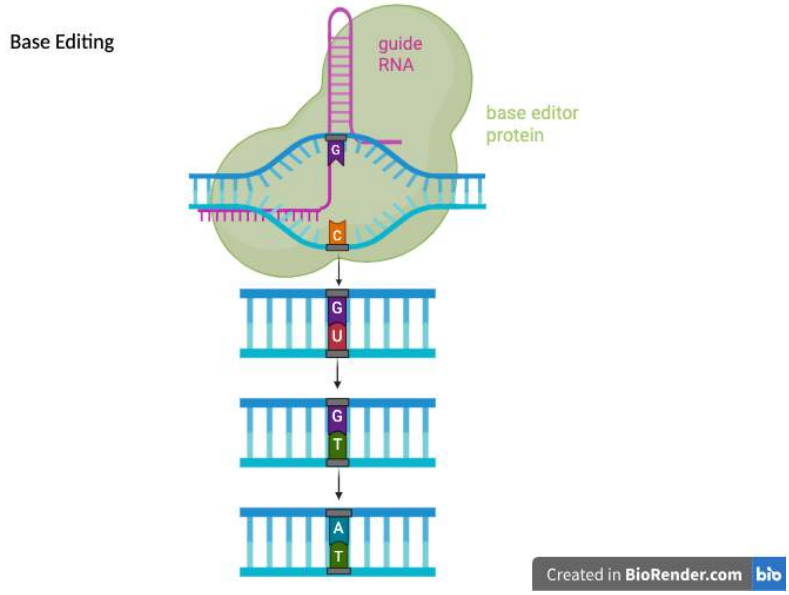


Figure 2. CRISPR Base Editing Mechanism.

Base editing is a CRISPR-derived genome editing approach that enables precise single-base substitutions without creating double-stranded DNA breaks or requiring a donor DNA template. In the illustrated cytosine base editing process, a cytosine (C) is deaminated to uracil (U), which is subsequently recognized by the cellular DNA repair machinery and converted into thymine (T), resulting in the conversion of a G:C base pair to an A:T base pair. *Created in BioRender. Zhou, E. (2026) https://app.biorender.com/illustrations/6948d9db5fb455d218ce65e8*

Prime editing is a more versatile technique that uses a modified Cas9 along with reverse transcriptase and a specialized guide RNA ("CRISPR Prime Editing Guide"). Prime editing can insert, delete, or replace DNA sequences with high precision and without creating double-strand breaks ("CRISPR Prime Editing Guide"). Although still less efficient than traditional CRISPR, prime editing represents a major advancement toward safer genetic therapies ("CRISPR Prime Editing Guide"). In this system, the modified Cas9 acts as a nickase that cuts only one strand of DNA, while the specialized guide RNA, called a prime editing guide RNA (pegRNA), both directs the enzyme to the target site and contains a template for the desired genetic change ("CRISPR Prime Editing Guide"). After the DNA strand is nicked, the pegRNA binds to the exposed strand, allowing the reverse transcriptase to synthesize new DNA that includes the intended edit ("CRISPR Prime Editing Guide"). The cell's repair machinery then uses this edited strand as a template to replace the original sequence, resulting in a precise and controlled genetic modification without the risks associated with double-strand breaks ("CRISPR Prime Editing Guide") (Figure 3).

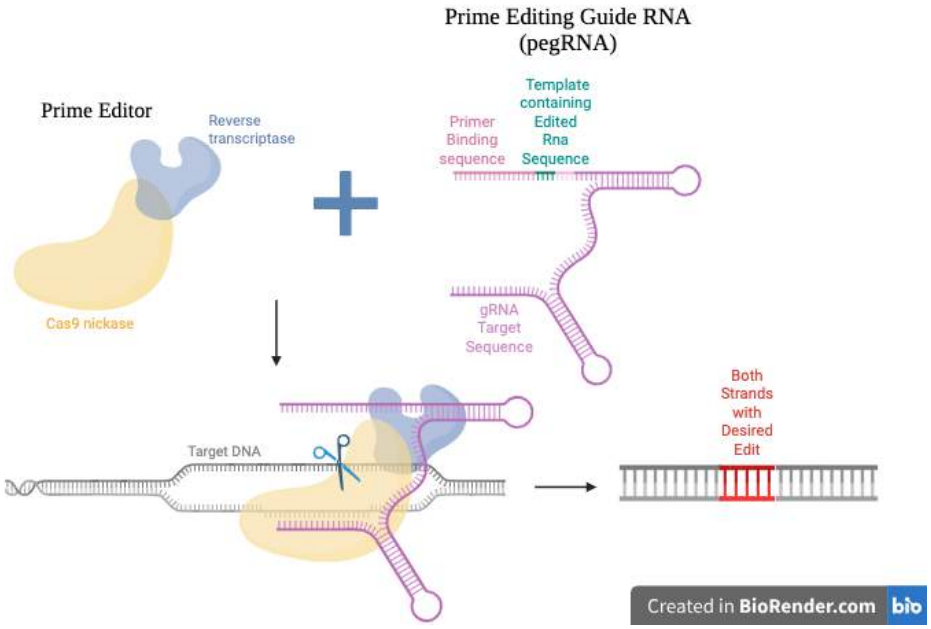


Figure 3. CRISPR Prime Editing Mechanism.

Prime editing employs a Cas9 nickase fused to a reverse transcriptase and a prime editing guide RNA (pegRNA) containing the target sequence, primer-binding site, and reverse transcription template encoding the desired edit. After the Cas9 nickase creates a single-strand break, the reverse transcriptase copies the edited sequence from the pegRNA into the target DNA. Cellular DNA repair processes then incorporate the edited sequence, resulting in precise base substitutions, insertions, or deletions without requiring double-stranded DNA breaks or donor DNA templates. *Created in BioRender. Zhou, E. (2026) https://app.biorender.com/illustrations/696ae747d0174125d5fa11f2*

Application of CRISPR

Herpes simplex keratitis is an eye disease caused by the herpes simplex virus (HSV), which can lead to blindness (Ahmad). Worldwide, nearly 491 million people between the ages of 15 and 49 are affected by HSV (Ahmad). Traditional antiviral drugs only suppress the virus but do not eliminate it from the body, because HSV can remain latent in nerve cells and reactivate later (Ahmad). Researchers have applied CRISPR knock-out to target and disrupt HSV genes, such as UL8 and UL29, which are necessary for viral replication (Ahmad). By knocking out these genes, the virus loses its ability to reproduce, reducing infection and preventing recurrent corneal damage (Ahmad). This is important because repeated HSV infections are a leading cause of blindness, and a genetic approach could provide lasting protection for patients' vision (Ahmad).

Hereditary tyrosinemia type I (HTI) is a rare genetic disorder caused by mutations in the FAH gene, which is required to break down the amino acid tyrosine (Wu et al.).Without a functional FAH gene, toxic byproducts build up in the liver and kidneys, which lead to organ damage (Wu et al.). Researchers have used CRISPR knock-in with HDR to correct the defective FAH gene in mouse models of HTI (Wu et al.). CRISPR knock-in is used because it allows precise correction of the mutation, making it easier to develop gene therapies for patients and restore normal enzyme function (Wu et al.). One method involves delivering CRISPR components through plasmids together with a DNA repair template, allowing the liver cells to repair the mutation and restore enzyme function (Wu et al.). In another approach, scientists transplanted hepatocytes that had been CRISPR-corrected in the lab back into diseased mice, successfully curing them (Wu et al.).

Future Direction of CRISPR

One of the most promising future directions of CRISPR technology is its role in personalized medicine (Wang et al.). Rather than using one standard treatment for all patients, CRISPR could allow therapies to be customized to an individual's unique genetic mutation (Wang et al.). Many genetic diseases are caused by specific DNA changes that vary from person to person, even within the same disorder (Wang et al.). By designing guide RNAs tailored to a patient's exact mutation, CRISPR-based therapies could precisely correct the faulty gene while leaving the rest of the genome unchanged ("CRISPR Methods"). This approach has the potential

to transform medicine from managing symptoms to permanently curing genetic conditions at their source (Carroll).

At the same time, the future success of CRISPR depends on ethical and controlled use (Liu et al.). While gene editing offers enormous medical benefits, it also raises serious ethical concerns, particularly regarding germline editing (Liu et al.). Germline editing refers to genetic modifications made to sperm, eggs, or embryos, which would pass genetic changes on to future generations (Gostimskaya). Without strict regulations, CRISPR could be misused for non-medical purposes such as genetic enhancement rather than disease treatment (Liu et al.). As CRISPR technologies continue to advance, scientists, policymakers, and ethicists must work together to establish clear guidelines that prioritize safety, fairness, and medical necessity (Liu et al.). Responsible oversight will ensure that CRISPR is used to heal rather than harm, allowing society to benefit from its power while avoiding ethical misuse (Liu et al.).

Overall, CRISPR technologies have significantly advanced the field of genome editing by providing researchers with increasingly precise and versatile tools for modifying DNA (Carroll). As discussed throughout this review, traditional CRISPR-Cas9, base editing, and prime editing each offer unique advantages for treating genetic diseases while presenting different technical limitations (Wang et al.). Continued improvements in editing accuracy, delivery methods, and safety will further expand the clinical applications of these technologies (Liu et al.). Although scientific and ethical challenges remain, ongoing research continues to move CRISPR closer to becoming a safe and effective platform for personalized medicine (Carroll). The future of CRISPR will be defined not only by what we can edit, but by what we choose to edit.

Works Cited

Ahmad, Irshad. “CRISPR Gene Therapy for Herpes Simplex Keratitis.” International Journal of Molecular Sciences, vol. 23, no. 20, 2022, PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC9569777/. Accessed 16 Feb. 2026.

Carroll, Dana. “Genome Editing: Past, Present, and Future.” Yale Journal of Biology and Medicine, vol. 90, no. 4, 2017, pp. 653–659. PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC5733845/. Accessed 23 Jan. 2026.

“CRISPR 101: Homology-Directed Repair.” Addgene Blog, Addgene, updated by Susanna Stroik, 24 Jan. 2023, https://blog.addgene.org/crispr-101-homology-directed-repair. Accessed 23 Jan. 2026.

“CRISPR Base Editing Guide.” Synthego, https://www.synthego.com/crispr-base-editing-guide. Accessed 23 Jan. 2026.

“CRISPR-Cas Gene Editing Teaching Resources.” Bio-Rad Laboratories, https://www.bio-rad.com/en-us/applications-technologies/crispr-cas-gene-editing-teaching-resources?ID=Q58I0DWDLBV5. Accessed 16 Feb. 2026.

“CRISPR Made Simple.” Innovative Genomics Institute, https://innovativegenomics.org/crispr-made-simple/. Accessed 16 Feb. 2026.

“CRISPR Methods.” Synthego, https://www.synthego.com/learn/crispr. Accessed 16 Feb. 2026.

“CRISPR Prime Editing Guide.” Synthego, https://www.synthego.com/guide/crispr-methods/prime-editing. Accessed 16 Feb. 2026.

“CRISPR Timeline.” Broad Institute, https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/crispr-timeline. Accessed 25 Feb. 2026.

Gostimskaya, Irina. “History and Development of CRISPR Genome Editing Technologies.” Journal of Genetic Engineering and Biotechnology, vol. 20, no. 1, 2022, PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC9377665/. Accessed 29 June 2026.

“History of Genome Engineering.” Synthego, https://www.synthego.com/learn/genome-engineering-history. Accessed 25 Feb. 2026.

Johannesen, Hana J., and Netanya Y. Spencer. “Overview: What Is CRISPR Screening?” Integrated DNA Technologies, 5 Aug. 2022, https://www.idtdna.com/pages/education/decoded/article/overview-what-is-crispr-screening. Accessed 25 Feb. 2026.

Liu, Wenyi, et al. “Limitations and Challenges of CRISPR-Based Genome Editing.” Frontiers in Cell and Developmental Biology, vol. 9, 2021, PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC8444435/. Accessed 29 June 2026.

Marnik, Elisabeth A., et al. “CRISPR: A New Way for Scientists to Edit DNA.” Frontiers for Young Minds, vol. 9, 2021, https://kids.frontiersin.org/articles/10.3389/frym.2021.600133. Accessed 25 Feb. 2026.

Ng, Daphne. “History of CRISPR.” Bitesize Bio, https://bitesizebio.com/47927/history-crispr/. Accessed 25 Feb. 2026.

Roberts, Rebecca. “Alternatives to CRISPR-Cas9 Nucleases for Next-Gen Therapy.” Synthego Blog, Synthego, https://www.synthego.com/blog/alternatives-to-crispr-cas9-nucleases-for-next-gen-therapy. Accessed 29 June 2026.

Wang, Si-Wei, et al. “CRISPR-Based Technologies for Genome Editing.” Molecular Cancer, vol. 21, no. 1, 2022, https://doi.org/10.1186/s12943-022-01518-8.

Wu, Shao-Shuai, et al. “CRISPR-Mediated Gene Correction for Hereditary Tyrosinemia Type I.” Signal Transduction and Targeted Therapy, vol. 5, 2020, PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC7150498/. Accessed 25 Feb. 2026.

Wyatt, David, and Dale Ramsden. “CRISPR 101: Non-Homologous End Joining.” Addgene Blog, Addgene, 1 Sept. 2020, https://blog.addgene.org/crispr-101-non-homologous-end-joining. Accessed 23 Jan. 2026.

Democracy, Human Rights, and Women's Political Representation as Pathways to Better Living Conditions After Independence By Neel Venkatesan

Abstract

Many post-colonial states achieved political independence yet continued to experience persistent poverty, low educational attainment, and high rates of preventable mortality. At the same time, some scholars argue that authoritarian governments can promote development more efficiently than democratic systems. This study examines whether countries that developed more democratic, rights-protecting, and politically inclusive institutions after entering the modern state system experienced greater improvements in population welfare after accounting for economic growth and time since independence.

Using the Correlates of War State List (2024), this study analyzes between 136 and 171 countries, depending on data availability for each outcome. Independent variables include mean democracy level (Our World in Data political regime index), change in the human rights index, mean women's parliamentary representation, change in GDP per capita, and years since state-system entry. Six welfare outcomes were analyzed using separate ordinary least squares (OLS) regression models with HC1 robust standard errors. Parallel models replacing democracy with the proportion of years spent under closed autocracy were estimated as a robustness comparison.

Countries experienced substantial average improvements in life expectancy, literacy, schooling, and mortality outcomes following state-system entry, although progress varied considerably across regions. After controlling for GDP growth and years since independence, improvements in human rights were significantly associated with gains in life expectancy ($\beta = 7.19$, $p < .05$) and reductions in infant mortality ($\beta = -3.39$, $p < .01$). Greater women's parliamentary representation was significantly associated with lower maternal mortality ($\beta = -4.56$, $p < .05$). Democracy exhibited statistically significant associations with five of six outcomes; however, negative coefficients for life expectancy and literacy appear consistent with ceiling effects, as supplementary level-based models produced positive associations for these outcomes. Closed-autocracy exposure showed comparatively weaker associations and reached statistical significance only for schooling.

Overall, the findings suggest that post-independence welfare trajectories are associated not only with economic development and time but also with institutional characteristics, particularly democracy, human rights and women's political representation. Because this study relies on observational cross-country data, these findings should be interpreted as associations rather than evidence of causal relationships.

1. Introduction

1.1 Background and Context

For much of modern history, European powers colonized Africa, Asia, the Caribbean, and the Pacific. These colonial powers extracted resources, restricted political authority, and imposed foreign rule (Acemoglu et al. 2001). The metrics most critical to human well-being—life expectancy, poverty rates, mortality rates, and literacy (United Nations; Roser et al. 2024)—were not prioritized by these colonial administrations (Acemoglu and Robinson 2012; Acemoglu et al. 2001). Post–World War II decolonization led to the creation of numerous new states; however, independence did not guarantee progress, as these nations inherited significant structural deficits from their colonizers (Acemoglu et al. 2001). The institutional trajectories that emerged after independence may therefore have been as consequential for long-term welfare as independence itself.

A substantial body of comparative research argues that long-run development is shaped by political and institutional arrangements (Acemoglu and Robinson 2012; Acemoglu et al. 2001). Democratic governance may broaden political accountability and improve population health (Roser and Ortiz-Ospina 2019), stronger human rights protections may improve access to public goods, and greater women's parliamentary representation may increase policy attention to health, education, and family welfare (Phillips 1995; Reynolds 2018; Coppedge et al. 2023). Together, these characteristics describe a model of inclusive governance. An alternative perspective argues that authoritarian governments can promote rapid development by concentrating political authority and insulating long-term investment decisions from electoral pressures (Przeworski and Limongi 1993; Gerring, Gjerløw, and Knutsen 2020). Whether more inclusive or more centralized institutional arrangements are more consistently associated with improvements in population welfare remains an empirical question (Gerring et al. 2012; Bollyky et al. 2019). This study addresses that question using cross-country evidence on six welfare outcomes measured since state-system entry (Singer et al. 2024; Roser et al. 2024).

1.2 Problem Statement and Rationale

Globally, significant disparities exist in life expectancy, poverty rates, schooling, and mortality rates among countries that gained independence from colonial rule (Roser et al. 2024; Acemoglu et al. 2001). These disparities likely reflect a combination of historical, economic, political, and institutional factors, making it difficult to isolate the contribution of any single dimension of governance. However, these gaps are not solely attributable to democracy, human rights, and women's representation in parliament (Gerring et al. 2012; Roser et al. 2024; Coppedge et al. 2023; Reynolds 2018). GDP differences and the time elapsed since independence also play a substantial role in a government's ability to enact improvements (World Bank; Singer et al. 2024). For example, a country with a higher GDP possesses more resources to invest in education, healthcare, and new technologies, while a country that has been independent for longer has had more time to establish infrastructure and strengthen governmental institutions. Therefore, any analysis of inclusive institutions must account for both GDP and time.

1.3 Significance and Purpose

This study contributes to ongoing debates over whether inclusive political institutions are associated with better long-term welfare outcomes than more authoritarian forms of governance (Acemoglu and Robinson 2012; Bollyky et al. 2019). By examining multiple dimensions of institutional development—including democracy, human rights protections, and women's political representation—while accounting for economic growth and time since independence (World Bank 2024; Singer et al. 2024), the analysis evaluates whether institutional characteristics remain associated with improvements in health, education, and poverty beyond broader patterns of economic development.

1.4 Objectives and Hypotheses

This study asks whether greater democratization (Benton et al. 2024), improvements in human rights (Roser et al. 2024; Coppedge et al. 2023), and greater women's parliamentary representation (Roser et al. 2024; Phillips 1995) are associated with larger improvements across six welfare outcomes than in countries with weaker inclusive institutions.

H1: Greater democratization will be associated with larger improvements in life expectancy, literacy, poverty reduction, infant mortality, maternal mortality, and schooling than less democratic or authoritarian governance, net of GDP growth and time elapsed.

H2: Larger improvements in human rights protections will be associated with greater gains across all six outcome measures than in countries where human rights protections weakened or stagnated, net of GDP growth and time elapsed.

H3: Greater women's parliamentary representation will be associated with larger improvements across all six outcome measures than in countries with lower female political inclusion, net of GDP growth and time elapsed.

Each hypothesis is evaluated against all six welfare outcomes.

Although these hypotheses anticipate positive associations between inclusive institutions and welfare outcomes, the analysis does not assume that institutional characteristics operate independently of broader historical or economic conditions (Acemoglu and Robinson 2012; Gerring et al. 2012). Instead, the empirical models estimate whether institutional variables remain associated with welfare outcomes after accounting for differences in GDP growth and years since state-system entry (World Bank 2024; Singer et al. 2024). Consequently, statistically significant coefficients should be interpreted as conditional associations rather than evidence of causal effects.

1.5 Scope and Limitations

The analysis includes only countries for which complete data were available (approximately 136–171 countries depending on the outcome) (Roser et al. 2024; World Bank 2024). Several potentially important determinants of development—including colonizer identity, duration of colonial rule, civil conflict, and women's economic participation—are not

incorporated into the models (Acemoglu et al. 2001; Phillips 1995). In addition, countries entered the state system with substantially different baseline levels of welfare, which may influence observed changes over time despite the supplementary level-based robustness analyses. Consequently, the findings should be interpreted as conditional associations rather than causal estimates. Finally, because comparable World Bank GDP data are largely available only after 2000 (World Bank 2024), sample sizes are smaller for models including GDP controls.

1.6 Theoretical Framework

This study draws primarily on institutional theories of development, which argue that the quality of political and economic institutions plays a central role in shaping long-term development outcomes (Acemoglu and Robinson 2012; Acemoglu et al. 2001). Inclusive institutions broaden political participation, strengthen accountability, and encourage investment in public goods and human capital, whereas extractive institutions concentrate political power and economic benefits among a relatively small elite. Following this framework, countries that developed more inclusive institutions after independence would be expected to experience greater improvements in population welfare than countries that maintained more extractive political systems.

An alternative perspective emphasizes the potential developmental advantages of centralized political authority. Proponents of the developmental-authoritarian model argue that governments insulated from electoral pressures may implement long-term investment strategies more efficiently than democratic governments, particularly during periods of industrialization or state building (Przeworski and Limongi 1993; Gerring, Gjerløw, and Knutsen 2020). Under this view, political concentration rather than political inclusion may facilitate economic development in low-income countries. The present study evaluates these competing perspectives by comparing democratic governance with sustained exposure to closed autocracy while examining multiple dimensions of human welfare.

The institutional variables included in this study represent complementary dimensions of inclusive governance. Mean democracy captures the extent to which political authority is broadly accountable to citizens (Benton et al. 2024). Changes in the human rights index measure the expansion or contraction of legal protections afforded to individuals (Coppedge et al. 2023; Roser et al. 2024). Mean women's parliamentary representation reflects the political inclusion of a historically underrepresented group and has been associated with increased legislative attention to health, education, and family welfare (Phillips 1995; Reynolds 2018). Rather than representing competing explanations, these variables capture distinct institutional characteristics that together describe the broader inclusiveness of political systems.

Economic development and historical context may also influence long-term welfare independently of political institutions. Wealthier countries generally possess greater fiscal capacity to invest in education, healthcare, and public infrastructure (World Bank 2024), while countries that have been independent longer have had more opportunities to build administrative capacity and expand public services (Singer et al. 2024; United Nations). Consequently, GDP

growth and years since state-system entry are included as control variables so that institutional associations can be evaluated after accounting for these broader structural influences.

The welfare outcomes examined in this study—life expectancy, literacy, poverty, infant mortality, maternal mortality, and mean years of schooling—represent widely recognized indicators of human well-being (Sen 1999; Roser et al. 2024). These outcomes also reflect areas in which colonial administrations frequently underinvested, making them appropriate indicators for evaluating post-independence institutional development (Acemoglu et al. 2001). If more inclusive institutions remain associated with improvements in these outcomes after accounting for economic growth and historical timing, the findings would provide evidence consistent with institutional theories of development. Conversely, if prolonged exposure to authoritarian governance exhibits stronger associations, the findings would lend greater support to developmental-authoritarian perspectives.

2. Methods

2.1 Research Design

This study employs a cross-sectional comparative research design in which each country contributes one observation for each welfare outcome. For every country, dependent variables were calculated as the absolute change between the first year with complete post-entry data and the most recent year with available observations (Singer et al. 2024; Roser et al. 2024). These change scores were then modeled as a function of long-term institutional characteristics following state-system entry.

The primary explanatory variables capture three complementary dimensions of institutional development: mean democracy, change in human rights protections, and mean women's parliamentary representation (Benton et al. 2024; Coppedge et al. 2023; Roser et al. 2024). To evaluate competing claims regarding authoritarian development, a parallel specification replaces mean democracy with the proportion of years each country was classified as a closed autocracy (OWID regime = 0). Democracy and closed-autocracy exposure were not included simultaneously because they represent opposing measures of the same underlying regime continuum and would introduce substantial multicollinearity.

Economic development and historical timing were incorporated as control variables through change in GDP per capita and years since state-system entry. These controls account for differences in fiscal capacity and the time available for institutional and social development that might otherwise confound institutional associations with welfare outcomes (World Bank 2024; Singer et al. 2024).

Each welfare outcome was estimated using a separate ordinary least squares (OLS) regression model with HC1 heteroskedasticity-robust standard errors. OLS was selected because the objective was to estimate average cross-country associations rather than predict individual country outcomes. Estimating separate models also accommodates differences in sample size

across dependent variables while allowing direct comparison of institutional coefficients across welfare measures.

2.2 Sample

The Correlates of War (COW) State System Membership List (Singer et al. 2024) provided the sampling frame for the analysis by identifying each country's entry into the modern international state system. Countries with multiple periods of statehood were assigned their most recent entry year to maintain consistency with contemporary institutional trajectories. Country names were subsequently harmonized with Our World in Data identifiers to facilitate merging across datasets. The analytic sample spans all major world regions, and entry dates range from the nineteenth century through the early 2000s, motivating the years-since-entry control.

Figure 1. Countries in the life-expectancy analytic sample (N = 171), colored by OWID world region.

Figure 2. Distribution of Correlates of War state-system entry decades (merged descriptive sample).

2.3 Variables and Measurements

Entry and present years: For each country and each outcome model, the entry year is the first year ≥ the COW entry year in which data are present for each variable (Singer et al. 2024). Present year = 2023 for all data unless a variable is unavailable; in that case present year = the most recent year before 2023 when all data are present (Roser et al. 2024; World Bank).

The primary independent variables measure long-run democratic exposure, institutional improvement in human rights, and women's parliamentary representation (Table 2.1) (Benton et al. 2024; Roser et al. 2024; Coppedge et al. 2023). In a parallel robustness specification, mean democracy is replaced by the share of post-entry years classified as closed autocracy (OWID regime = 0) (Benton et al. 2024); democracy and closed-autocracy exposure are not entered in the same model.

Institutional variables were selected because they represent complementary dimensions of political inclusiveness rather than alternative measures of the same concept.

Table 2.1. Independent variables

Variable	Operational definition	Source	Ref.
Mean democracy index	Average OWID political regime score (0–3) from merge year through present year	OWID, political-regime.csv	Benton et al. 2024
Human rights index change	V-Dem human rights index at present minus entry (0–1 scale)	OWID, human-rights-index-vdem.csv	Roser et al. 2024; Coppedge et al. 2023
Mean women in parliament share	Average percentage of lower-chamber seats held by women, merge year to present	OWID, share-of-women-in-parliament.csv	Roser et al. 2024
Share of years at closed autocracy	Share of years at regime = 0; parallel Table 3 models only	OWID, political-regime.csv	Benton et al. 2024

Control Variables

GDP per capita change and years since state-system entry were included as controls because both are expected to influence long-term welfare independently of political institutions (World Bank 2024; Singer et al. 2024).

Table 2.2 Control Variables

Variable	Operational definition	Source	Ref.
GDP per capita change	Present minus entry GDP per capita (constant 2017 intl. $, PPP)	World Bank, gdp-per-capita-worldbank.csv	World Bank
Years since entry	Present year minus merge year	Derived from COW + OWID	Singer et al. 2024

Dependent Variables (changes, present − entry)

Dependent variables are absolute changes in population welfare outcomes from the entry-baseline year to the present year (present − entry; Table 2.3) (Roser et al. 2024; United Nations). Because lower poverty and mortality represent improved welfare, negative change values correspond to favorable outcomes for these variables.

Table 2.3. Dependent variables

Variable	Operational definition	Source	Ref.
Life expectancy change	Change in life expectancy at birth (years)	OWID, life_expentancy.csv	Roser et al. 2024
Literacy change	Change in adult literacy rate (percentage points)	OWID, cross-country-literacy-rates.csv	Roser et al. 2024
Extreme poverty change	Change in share below $3/day (percentage points)	OWID, share-of-population-in-extreme-poverty.csv	Roser et al. 2024
Infant mortality change	Change in infant deaths per 1,000 live births	OWID, infant-mortality.csv	Roser et al. 2024
Maternal mortality change	Change in maternal deaths per 100,000 live births	OWID, maternal-mortality.csv	Roser et al. 2024
Mean years of schooling change	Change in mean years of schooling (ages 15–64)	OWID, mean-years-of-schooling-long-run.csv	Roser et al. 2024

2.4 Procedure

The analytical workflow consisted of five steps. First, the Correlates of War State List was used to identify each country's state-system entry year. Second, countries were harmonized across datasets using standardized OWID identifiers. Third, institutional variables, welfare outcomes, GDP change, and years since entry were calculated for each country. Fourth, observations with incomplete data for any variable included in a given model were excluded. Finally, separate OLS regression models were estimated for each welfare outcome under the primary democracy specification, the alternative closed-autocracy specification, and corresponding univariate models.

2.5 Data Analysis

For each dependent variable Y, the estimated model is:

$$\Delta Y\_i = \beta_0 + \beta_1 MeanDemocracy\_i + \beta_2 \Delta HumanRights\_i + \beta_3 MeanWomenParl\_i + \beta_4 \Delta GDPpc\_i + \beta_5 YearsSinceEntry\_i + \varepsilon\_i$$

Separate regression models were estimated for each dependent variable because data availability differed across welfare indicators, resulting in varying analytic samples. Estimating each equation independently ensured that coefficients were derived from the maximum number of available observations for each outcome. HC1 robust standard errors were used to reduce sensitivity to heteroskedasticity, allowing statistical inference without assuming constant error variance across countries.

Six dependent variables were analyzed. The primary model used mean democracy; a parallel specification replaced it with the share of years at closed autocracy (regime = 0), holding other predictors constant. Models were fitted by OLS with HC1 robust standard errors using Python (statsmodels). Significance was assessed at $\alpha = 0.10$, 0.05, and 0.01. OWID and World Bank coverage determine sample size (~136–171 countries per outcome). The subscript i indexes countries in the cross-country sample (one observation per country per outcome). Y is the welfare outcome being modeled, and ΔY_i is the change in that outcome for country i from its state-system entry year to the present year (present value minus entry value). As a robustness check on ceiling effects, each Table 1 model was re-estimated with the dependent variable replaced by the post-entry mean outcome level (average from merge year through present), holding predictors unchanged.

Although several welfare outcomes are bounded or exhibit skewed distributions, OLS remains appropriate for estimating conditional mean associations in cross-country comparative analyses of this size and provides coefficients that are readily comparable across models.

2.6 Ethical Considerations

The study uses publicly available, aggregated secondary data with no human subjects.

3. Results

3.1 Descriptive Statistics and Model-Free Evidence

Descriptive statistics were calculated separately for each analytic sample because the inclusion of GDP and complete outcome series reduced sample sizes from 171 countries to between 136 and 171 observations, depending on the outcome. Across the analytic samples, countries experienced substantial improvements in population welfare following state-system entry. On average, life expectancy increased by approximately 8.9 years, literacy increased by 8.2 percentage points, and mean years of schooling increased by nearly 2.5 years (Roser et al. 2024). Infant mortality, maternal mortality, and extreme poverty generally declined over the same period, although the magnitude of improvement varied considerably across countries. Mean women's parliamentary representation also increased substantially, averaging approximately 23 percentage points in the life-expectancy sample (Roser et al. 2024).

These descriptive statistics demonstrate considerable heterogeneity in post-independence development trajectories. Because countries differed substantially in both institutional development and welfare improvements, multivariate regression models were estimated to examine whether institutional characteristics remained associated with these outcomes after accounting for economic growth and years since state-system entry.

Figure 3. Mean life-expectancy change since state-system entry by OWID world region.

Welfare gains are widespread albeit uneven: life expectancy and literacy changes are right-skewed, while infant and maternal mortality changes cluster toward large improvements (negative values) (Roser et al. 2024). The distributions shown in Figure 4 indicate substantial variation across countries in each welfare outcome. Life expectancy and literacy changes exhibit moderate right skew, whereas improvements in infant and maternal mortality are concentrated among larger negative changes (indicating greater reductions in mortality). This variation supports further examination of institutional and economic correlates of post-independence welfare outcomes.

The bivariate correlations in the merged descriptive sample show expected results (Figure 5). Gains in life expectancy and literacy move together strongly ($r = 0.67$), and countries that democratized more also tended to see larger life expectancy gains ($r = 0.61$) and larger poverty reductions ($r = 0.68$) (Roser et al. 2024; Benton et al. 2024). Countries with larger infant mortality declines also registered larger gains in life expectancy ($r = 0.70$). Consistent with expectations, countries experiencing larger declines in infant mortality also tended to experience larger gains in life expectancy ($r = 0.70$), reflecting the close relationship between improvements in child survival and overall population longevity. These patterns suggest that development and inclusive institutions move forward together in the post-decolonization era, motivating multivariate models that hold GDP and time constant (World Bank; Singer et al. 2024).

Bivariate correlations alone, however, cannot distinguish the role of institutions from the effects of economic growth and time (World Bank; Singer et al. 2024). A country that democratized may also have grown wealthier over the same period, and a country that has been independent longer has had more time to improve on all these outcomes regardless of how it is governed. The multivariate models in the next section are designed to separate those effects,

isolating the association between institutional characteristics and welfare outcomes after the contributions of wealth and time are accounted for.

3.2 Regression Results

The empirical analyses indicate that the three institutional variables are associated with different dimensions of post-independence welfare. After accounting for GDP growth and years since state-system entry, improvements in human rights exhibited the most consistent associations with health outcomes, women's parliamentary representation was most strongly associated with maternal mortality, and mean democracy was associated with five of the six welfare outcomes examined. Results from the alternative closed-autocracy specification were generally weaker.

Improvements in the human rights index were positively associated with gains in life expectancy ($\beta = 7.19$, $p < .05$) and negatively associated with infant mortality ($\beta = -3.39$, $p < .01$) after controlling for GDP growth and years since state-system entry. No statistically significant associations were observed for literacy, poverty, maternal mortality, or schooling. Among the institutional variables examined, improvements in human rights exhibited the most consistent associations with health-related outcomes.

Greater women's parliamentary representation was significantly associated with reductions in maternal mortality ($\beta = -4.56$, $p < .05$). Associations with the remaining welfare outcomes were not statistically significant.

Mean democracy was negatively associated with changes in life expectancy and literacy, and positively associated with changes in poverty and mortality. Because larger reductions in poverty and mortality correspond to more negative change scores, the positive coefficients on these outcomes indicate that, in raw change terms, more democratic countries experienced smaller reductions in poverty, infant mortality, and maternal mortality, not larger ones. Because these outcomes were measured as absolute changes from state-system entry, these patterns do not necessarily imply poorer welfare outcomes among more democratic countries. One possible explanation is that countries with higher average democracy entered the observation period with relatively favorable baseline conditions, leaving less room for subsequent improvement. This interpretation is evaluated using supplementary level-based models presented in Table 6, which show a full sign reversal consistent with a ceiling effect for life expectancy, literacy, poverty, and infant mortality. The reversal is not statistically significant for maternal mortality, so the ceiling-effect explanation is better supported for the other four outcomes than for maternal mortality.

GDP per capita change and years since state-system entry were statistically significant in most models, indicating that economic development and historical timing remained important correlates of welfare improvement. Nevertheless, institutional variables retained statistically significant associations with several outcomes after these controls were included.

Table 1. Association of democracy, institutional change, and economic change with population outcome change since state-system entry (OLS, HC1 robust standard errors)

Independent variable	(1) Life expectancy Δ (years)	(2) Literacy Δ (pp)	(3) Extreme poverty Δ (pp)	(4) Infant mortality Δ (per 1,000)	(5) Maternal mortality Δ (per 100k)	(6) Schooling Δ (years)
Mean democracy (0–3)	−2.08*** *(0.53)*	−1.50* *(0.78)*	7.37*** *(2.20)*	1.15*** *(0.20)*	41.11** *(20.97)*	−0.14 *(0.11)*
Human rights Δ	7.19** *(2.92)*	−1.41 *(5.95)*	9.44 *(13.37)*	−3.39*** *(1.20)*	−42.22 *(107.44)*	0.03 *(0.48)*
Women in parliament (%) (mean share)	0.09* *(0.05)*	−0.05 *(0.06)*	−0.00 *(0.13)*	−0.03 *(0.02)*	−4.56** *(2.12)*	0.00 *(0.01)*
GDP per capita Δ (intl. $)	−0.000011 *(0.000020)*	−0.000209*** *(0.000080)*	0.000420** *(0.000179)*	0.000034* *(0.000018)*	0.003739** *(0.001539)*	0.000019 *(0.000012)*
Years since entry	0.33*** *(0.04)*	0.59*** *(0.07)*	−0.83*** *(0.15)*	−0.16*** *(0.03)*	2.06 *(2.92)*	0.12*** *(0.01)*
Constant	0.10 *(1.18)*	1.46 *(1.54)*	−12.42*** *(3.69)*	0.36 *(0.90)*	−164.18*** *(39.66)*	−0.39 *(0.27)*
Observations	171	136	160	169	169	165
R^2	0.238	0.366	0.243	0.363	0.194	0.357
Adjusted R^2	0.215	0.341	0.219	0.344	0.169	0.337

Notes: pp = percentage points. Mean democracy = average OWID political regime index (0–3) from merge year to present. Mean women in parliament = average lower-chamber share (%) over the same window. GDP per capita from the World Bank (international $). For poverty and mortality outcomes, a negative coefficient means larger values of the predictor are associated with larger reductions in the outcome. *** p < 0.01, ** p < 0.05, * p < 0.10.

To evaluate whether the democracy coefficients reflected ceiling effects rather than poorer welfare outcomes, Table 6 re-estimated the primary models using post-entry mean outcome levels instead of absolute change scores.

Table 2. Mean democracy coefficients under change and post-entry mean-level specifications (same controls as Table 1; OLS, HC1 robust standard errors)

Outcome	Mean democracy coefficient on change (present-entry)	Mean democracy coefficient on post-entry level mean	N
Life expectancy	-2.08*** (0.53)	3.72*** (0.75)	171
Literacy	-1.50* (0.78)	3.49* (1.91)	136
Extreme Poverty	7.37*** (2.20)	-5.61** (2.30)	160
Infant Mortality	1.15*** (0.20)	-1.32*** (0.22)	169
Maternal Mortality	41.11** (20.97)	-41.46 (31.58)	169
Mean Years of Schooling	-0.14 (0.11)	1.20*** (0.26)	165

Note: Change models use absolute welfare change (present − entry), as in Table 1. Level models replace the dependent variable with the average post-entry outcome level (merge year through present), parallel to mean democracy and mean women's parliamentary share. All models include human rights change, mean women in parliament, GDP per capita change, and years since entry. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Across several outcomes, the sign of the democracy coefficient reversed when post-entry mean levels rather than absolute change scores were modeled. For life expectancy, literacy, and schooling, higher mean democracy was positively associated with average post-entry welfare levels despite being negatively associated with change scores. Similar reversals occurred for poverty and infant mortality, where lower values represent more favorable outcomes. These findings are consistent with the possibility that countries with higher average democracy entered the observation period with relatively favorable baseline conditions, resulting in smaller absolute improvements over time.

3.2.1 Authoritarian Regime Comparison

To compare democratic and authoritarian exposure on the same scale, the mean democracy index was replaced with the share of years at closed autocracy (OWID regime = 0) (Benton et al. 2024). Human rights change (Roser et al. 2024; Coppedge et al. 2023), mean women's parliamentary share (Roser et al. 2024), GDP per capita change (World Bank 2024), and years since entry (Singer et al. 2024) were unchanged. Democracy and closed-autocracy share mirror images of the same regime series and were not entered together. Table 3 reports the authoritarian specification. Mean democracy was significant for five of six outcomes in Table 1 (all except schooling) (Benton et al. 2024). Closed-autocracy exposure was significantly associated only with schooling ($\beta = 0.71$, $p < .05$). No statistically significant associations were observed for the remaining welfare outcomes. Human rights improvements remained significantly associated with gains in life expectancy and reductions in infant mortality, consistent with the primary specification. Overall model fit was modestly lower than in the democracy models.

Table 3. Parallel multivariate models replacing mean democracy with share of years at closed autocracy (OLS, HC1 robust standard errors; other predictors identical to Table 1)

Independent variable	(1) Life expectancy Δ (years)	(2) Literacy Δ (pp)	(3) Extreme poverty Δ (pp)	(4) Infant mortality Δ (per 1,000)	(5) Maternal mortality Δ (per 100k)	(6) Schooling Δ (years)
Closed autocracy share (regime = 0)	3.07 *(2.07)*	4.01 *(2.52)*	−8.39 *(8.68)*	−0.17 *(0.68)*	28.20 *(42.18)*	0.71** *(0.34)*
Human rights Δ	7.61** *(3.01)*	−1.04 *(6.06)*	9.92 *(13.60)*	−3.83*** *(1.23)*	−39.99 *(109.72)*	0.08 *(0.47)*

Women in parliament (%) (mean share)	0.04 *(0.05)*	−0.06 *(0.06)*	0.11 *(0.13)*	0.01 *(0.02)*	−4.10* *(2.20)*	−0.00 *(0.01)*
GDP per capita Δ (intl. $)	−0.000054** *(0.000027)*	−0.000212** *(0.000088)*	0.000682*** *(0.000201)*	0.000069*** *(0.000019)*	0.004524*** *(0.001650)*	0.000019* *(0.000011)*
Years since entry	0.29*** *(0.04)*	0.59*** *(0.07)*	−0.81*** *(0.16)*	−0.13*** *(0.03)*	6.13** *(2.39)*	0.12*** *(0.01)*
Constant	−1.17 *(1.61)*	−1.01 *(1.14)*	−3.72 *(3.75)*	0.41 *(0.91)*	−190.50*** *(43.34)*	−0.70** *(0.27)*
Observations	171	136	160	169	169	165
R^2	0.181	0.366	0.180	0.256	0.173	0.371
Adjusted R^2	0.156	0.342	0.153	0.234	0.147	0.351

Note: Closed autocracy = OWID regime category 0. *** p < 0.01, ** p < 0.05, * p < 0.10.

3.2.2 Post-1945 State-System Entry Subsample

As a robustness check, Table 5 re-estimates the Table 1 models for countries that entered the Correlates of War state system in 1945 or later (N = 98–119 per outcome) (Singer et al. 2024). Estimated coefficients were broadly consistent with those reported in the primary analysis. Human rights improvements remained significantly associated with life expectancy and infant mortality, while women's parliamentary representation remained significantly associated with maternal mortality. These results indicate that the principal findings were similar when the sample was restricted to countries entering the state system in 1945 or later.

Table 4. Multivariate OLS coefficients for countries with state-system entry in 1945 or later (same specification as Table 1; HC1 robust standard errors)

Independent variable	**(1) Life expectancy Δ (years)**	**(2) Literacy Δ (pp)**	**(3) Extreme poverty Δ (pp)**	**(4) Infant mortality Δ (per 1,000)**	**(5) Maternal mortality Δ (per 100k)**	**(6) Schooling Δ (years)**
Mean democracy (0–3)	−1.93*** *(0.65)*	0.59 *(1.07)*	3.65* *(2.03)*	1.01*** *(0.24)*	24.26 *(25.06)*	−0.12 *(0.14)*
Human rights Δ	10.08*** *(3.60)*	−6.35 *(7.54)*	18.38 *(17.75)*	−4.20*** *(1.55)*	−31.55 *(155.75)*	0.24 *(0.74)*
Women in parliament (%) (mean share)	0.15** *(0.06)*	−0.05 *(0.07)*	0.05 *(0.16)*	−0.05** *(0.03)*	−6.24** *(2.80)*	0.00 *(0.01)*
GDP per capita Δ (intl. $)	−0.000014 *(0.000027)*	−0.000181*** *(0.000062)*	0.000786*** *(0.000240)*	0.000040 *(0.000025)*	0.007310*** *(0.002454)*	0.000027 *(0.000017)*
Years since entry	0.31*** *(0.05)*	0.67*** *(0.08)*	−1.03*** *(0.19)*	−0.16*** *(0.03)*	0.55 *(4.71)*	0.12*** *(0.01)*

Constant	−0.49 *(1.34)*	−1.99 *(2.05)*	−6.90 *(4.57)*	0.94 *(0.88)*	−118.31** *(59.81)*	−0.46 *(0.32)*
Observations	119	98	112	117	117	113
R^2	0.246	0.418	0.255	0.351	0.207	0.402
Adjusted R^2	0.213	0.387	0.220	0.322	0.171	0.374

Note: Sample restricted to COW state-system entry year ≥ 1945. *** p < 0.01, ** p < 0.05, * p < 0.10.

3.3 Summary of Hypothesis Tests

Table 5. Summary of hypothesis test results

Hypothesis	Life expectancy	Literacy	Poverty	Infant mortality	Maternal mortality	Schooling
H1 (democracy)	Yes*** (a)	Yes* (a)	Yes*** (a)	Yes*** (a)	Yes** (b)	No
H2 (human rights)	Yes**	No	No	Yes***	No	No
H3 (women in parliament)	Partial (p ≈ 0.10)	No	No	No	Yes**	No

Note: H1 for life expectancy, literacy, poverty, and infant mortality is interpreted with Table 6 level models because change coefficients reflect ceiling effects. The maternal mortality result is tentative because the level-based model is not significant.

4. Discussion

4.1 Restatement of Key Findings

The findings indicate that no single institutional characteristic was consistently associated with improvements across all six welfare outcomes. Instead, different dimensions of institutional development exhibited distinct patterns of association. Improvements in human rights were most consistently associated with health-related outcomes, particularly gains in life expectancy and reductions in infant mortality, whereas greater women's parliamentary representation was most strongly associated with lower maternal mortality. These associations remained statistically significant after accounting for GDP growth and years after state-system entry (World Bank 2024; Singer et al. 2024), suggesting that institutional characteristics may be associated with welfare outcomes beyond differences in economic development alone.

The results also highlight the importance of interpreting democracy measures carefully. Mean democracy was significantly associated with five of the six outcomes examined, but several coefficients were negative when welfare was measured as absolute change from state-system entry. Supplementary level-based models (Table 6) indicate that these patterns are consistent with ceiling effects rather than poorer welfare outcomes among more democratic countries. Countries with higher average democracy generally exhibited more favorable average welfare levels despite experiencing smaller absolute improvements over time, likely reflecting more advantageous starting conditions (Acemoglu and Robinson 2012). Consequently, the

findings on the effects of democracy on change in welfare outcomes should not be interpreted as evidence that democratic governance reduces welfare, but rather that baseline conditions influence the magnitude of subsequent change.

Together, these results suggest that institutional development is multidimensional. Democracy, human rights protections, and women's political representation appear to capture different aspects of governance and are associated with different dimensions of population welfare rather than operating as interchangeable measures of institutional quality.

4.2 Implications for Post-Colonial Development

The findings contribute to ongoing debates regarding the relationship between governance and development in post-colonial states. Economic growth and time since independence remained important correlates of welfare improvement, yet institutional variables retained statistically significant associations with several health outcomes after these factors were controlled. In particular, improvements in human rights were consistently associated with gains in life expectancy and reductions in infant mortality across both the primary and alternative model specifications (Roser et al. 2024; Coppedge et al. 2023).

These findings do not imply that strengthening human rights protections alone will improve population welfare. Rather, they suggest that countries exhibiting greater progress in human rights also tended to experience better health outcomes than would be expected based on economic growth and historical timing alone. Similarly, greater women's parliamentary representation was associated with lower maternal mortality, consistent with previous research linking women's political power to maternal-health outcomes and broader policy attention toward health and family welfare (Bhalotra et al. 2023; Phillips 1995; Reynolds 2018). Although causal relationships cannot be established, the results support continued investigation into the institutional dimensions of post-colonial development alongside more traditional economic explanations.

4.3 Connection to Objectives

The study's objectives were to evaluate whether democracy, improvements in human rights, and women's parliamentary representation were associated with long-term improvements in post-colonial welfare after accounting for economic development and time since independence. Overall, the evidence provides partial support for each hypothesis. Human rights improvements were consistently associated with gains in life expectancy and reductions in infant mortality. Women's parliamentary representation was significantly associated with lower maternal mortality. Democracy exhibited statistically significant associations across several outcomes, although the direction of some coefficients reflected the influence of baseline conditions rather than straightforward improvements in welfare. Collectively, these findings support the broader proposition that multiple dimensions of institutional development are associated with population welfare, while also demonstrating that no single institutional characteristic explains every outcome examined.

4.4 Limitations

Several limitations should be considered when interpreting these findings. First, the study is observational and cross-sectional, meaning statistically significant associations should not be interpreted as evidence of causal relationships. Unobserved factors—including colonial administrative strategies, civil conflict, geography, state capacity, and public investment in health and education—may influence both institutional development and welfare outcomes (Acemoglu et al. 2001). The influence of colonial legacies on institutional development may also fade over time (Maseland 2018), which the years-since-entry control only partially captures.

Second, countries entered the international state system with substantially different baseline levels of development. Although supplementary level-based analyses were conducted to evaluate potential ceiling effects, future research could model baseline outcome levels directly or employ longitudinal panel designs to better distinguish institutional effects from long-term convergence.

Third, GDP data availability reduced sample sizes for several models, and women's parliamentary representation represents only one dimension of broader gender inclusion. Finally, the analysis relies on national averages, which may conceal important within-country variation in welfare outcomes and institutional quality.

4.5 Recommendations for Future Research

Future research could strengthen these findings through longitudinal panel analyses with country fixed effects, allowing institutional changes and welfare outcomes to be examined within countries over time rather than across countries alone. Incorporating colonial administrative characteristics, conflict exposure, public health expenditure, educational investment, and additional measures of gender inclusion would also help clarify the mechanisms linking institutions and development. Finally, comparative case studies of countries that experienced unusually rapid welfare improvements could complement the cross-country analysis by identifying institutional pathways that are difficult to capture using aggregate statistical models.

4.6 Closing Thought

This study examined whether institutional development following independence was associated with long-term improvements in population welfare across post-colonial states. Although economic growth and time since independence remained important correlates of development, institutional characteristics also exhibited meaningful associations with several welfare outcomes. Improvements in human rights were consistently associated with gains in life expectancy and reductions in infant mortality, while greater women's parliamentary representation was associated with lower maternal mortality. Democracy demonstrated broader but more variable associations, with supplementary analyses suggesting that several counterintuitive coefficients reflected differences in baseline conditions rather than poorer welfare outcomes.

Works Cited

Acemoglu, Daron, and James A. Robinson. Why Nations Fail: The Origins of Power, Prosperity, and Poverty. Crown Business, 2012.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. "The Colonial Origins of Comparative Development: An Empirical Investigation." American Economic Review, vol. 91, 2001, pp. 1369-1401.

Benton, Joe, et al. "Political Regime." Our World in Data, 2024, ourworldindata.org/grapher/political-regime.

Bhalotra, Sonia, et al. "Maternal Mortality and Women's Political Power." Journal of the European Economic Association, vol. 21, no. 4, 2023, pp. 1715-54.

Bollyky, Thomas J., et al. "The Relationships between Democratic Experience, Adult Health, and Cause-Specific Mortality in 170 Countries between 1980 and 2016: An Observational Analysis." The Lancet, vol. 393, no. 10181, 2019, pp. 1628-40.

Coppedge, Michael, et al. "V-Dem Country-Year Dataset v13." Varieties of Democracy Project, 2023, www.v-dem.net.

Gerring, John, Strom C. Thacker, and Rodrigo Alfaro. "Democracy and Human Development." The Journal of Politics, vol. 74, no. 1, 2012, pp. 1-17.

Maseland, Robbert. "Is Colonialism History? The Declining Impact of Colonial Legacies on African Institutional and Economic Development." Journal of Institutional Economics, vol. 14, no. 2, 2018, pp. 259-87.

Phillips, Anne. The Politics of Presence. Oxford University Press, 1995.

Reynolds, Andrew. "Gender and the Politics of Death: Female Representation, Political and Developmental Context, and Population Health in a Cross-National Panel." Demography, vol. 55, no. 5, 2018, pp. 1825-49.

Roser, Max, and Esteban Ortiz-Ospina. "Does Democracy Lead to Better Health?" Our World in Data, 2019, ourworldindata.org/democracy-health.

Roser, Max, et al. "Human Rights Index (V-Dem)." Our World in Data, 2024, ourworldindata.org/grapher/human-rights-index-vdem.

Roser, Max, et al. "Life Expectancy." Our World in Data, 2024, ourworldindata.org/life-expectancy.

Roser, Max, et al. "Literacy Rate." Our World in Data, 2024, ourworldindata.org/literacy.

Roser, Max, et al. "Share of Population in Extreme Poverty." Our World in Data, 2024, ourworldindata.org/extreme-poverty.

Roser, Max, et al. "Share of Women in Parliament." Our World in Data, 2024, ourworldindata.org/grapher/share-of-women-in-parliament.

Sen, Amartya. Development as Freedom. Oxford UP, 1999.

Singer, David, et al. "Correlates of War Project International System Membership Codebook, 1816-2024." Correlates of War Project, 2024, correlatesofwar.org.

United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations, 2015, sdgs.un.org/2030agenda.

World Bank. "GDP per Capita, PPP (Constant 2017 International $)." World Development Indicators, 2024, data.worldbank.org.

Art Therapy as a Non-Pharmacological Intervention: A Creative Bridge between Psychological and Neurocognitive Recovery By Sofia Neubauer[1]

Abstract

Recovery from illnesses or traumatic injuries is a complex, physically and emotionally draining process. For patients navigating oncology treatments or traumatic brain injury (TBI), the healing journey is often disrupted by psychological distress, identity loss, and neurocognitive impairment. Conventional pharmacological approaches often target isolated symptoms rather than the integrated emotional and neural systems required for holistic healing. As healthcare increasingly shifts toward interdisciplinary models of care, there is a growing need for non-pharmacological interventions that can simultaneously address emotional regulation and brain plasticity in these clinical contexts.

This research explores the effectiveness of visual art therapy—the guided use of painting, drawing, or sculpting within a therapeutic relationship—as a multi-system intervention for psychological and neurocognitive recovery. Art therapy functions as a space where patients can externalize non-verbal emotions, reclaiming a sense of agency and reconstructing a coherent identity. This paper analyzes how creative engagement facilitates recovery through three mechanisms: emotional regulation, identity reconstruction, and neurocognitive engagement. Among oncology patients, art therapy is shown to lower physiological indicators of stress and reduce anxiety and depressive symptoms. For individuals recovering from TBI, the process supports the integration of pre-illness and post-illness narratives. Furthermore, neurocognitive research indicates that art-making engages cross-hemispheric communication and enhances functional connectivity within neural networks responsible for attention, memory, and executive functioning.

Keywords art therapy, emotional regulation, identity reconstruction, neurocognitive engagement, neuroplasticity

Research Question: How does engagement in visual art therapy influence psychological and neurocognitive recovery outcomes across clinical populations, particularly patients with cancer and individuals recovering from traumatic brain injury?

1. Introduction

1.1 Non-pharmacological Interventions and Art Therapy

Non-pharmacological interventions are increasingly relevant in contemporary healthcare as clinicians recognize that psychological well-being plays a crucial role in health outcomes, influencing recovery speed, treatment adherence, and overall quality of life. Particularly among children and adolescents with long-term illnesses, the treatment trajectory can become life-altering and traumatic, disrupting identity formation, peer relationships, and emotional

development (Manning et al.). Prolonged illness can contribute to an overall loss of personal identity, increased social isolation, and extreme confusion and uncertainty (Manning et al.).

Patients coping with chronic medical or neurological conditions frequently experience anxiety, depression, or emotional distress following an injury or illness, especially when their condition diminishes their sense of control. This loss of control may intensify identity loss, whereas restoring a sense of agency can positively impact health outcomes (Yeom and Lee). Despite these insights, traditional pharmacological approaches, which primarily target symptom reduction, frequently fail to address the emotional aspects of recovery.

As a result, non-pharmacological interventions such as art therapy have gained increasing attention for their association with enhanced emotional regulation, increased autonomy, and improved cognitive and sensory integration (Martinelli et al.). Specifically, Martinelli and colleagues note that Active Visual Art Therapy (AVAT) is effective in reducing pain, anxiety, and depressive symptoms in pediatric and adult patients while supporting autonomy and beneficence. Importantly, art therapy interventions in hospitals are often low-maintenance and low-cost relative to many medical interventions, while still yielding meaningful psychological and cognitive benefits (Martinelli et al.).

Art therapy is a psychotherapeutic approach that utilizes creative art-making under the guidance of a trained therapist to improve mental, emotional, and physical well-being (Joschko et al.). Its wide range of methods and media, most commonly painting, drawing, and sculpting, makes it highly accessible. Patients undergoing extensive treatments may struggle with emotional regulation, defined as the ability to recognize, understand, and manage emotional responses (McRae and Gross). Psychological distress, a state of emotional suffering characterized by symptoms of depression, anxiety, and impaired coping, is common in chronically ill patients (Burnette et al.). Such distress is frequently accompanied by reductions in quality of life, a multidimensional measure encompassing an individual's physical health, psychological state, level of independence, and emotional coping abilities (McPherson et al.).

Despite growing interest, the study of art therapy and its effects has proven methodologically complex. Many studies rely on small sample sizes, which can limit generalizability. Intervention protocols and outcome measures vary widely, making cross-study comparison difficult. Isolating the psychological mechanisms driving therapeutic outcomes—such as whether improvements stem from creative expression, social interaction, or neural remodeling—remains challenging. While the use of art therapy in these populations has expanded and results are promising, the current body of high-quality evidence remains insufficient (Iguina and Kashan). These challenges highlight the need for standardized, interdisciplinary research to clarify how art therapy contributes to mental and neurological recovery.

2. Analytical Framework

2.1 Analytical Focus

This paper adopts an analytical framework that explores the significance of art therapy as a bridge between psychological and neurocognitive recovery. By analyzing existing research and clinical evidence, it clarifies how creative expression facilitates healing, not just as an emotional release, but as a mechanism that influences neural processes and overall well-being. The analysis focuses on the seamless integration of both psychological and neurological processes.

On the psychological level, the analysis investigates how structured creative engagement in the arts supports emotional regulation, reduces anxiety, and enhances coping strategies. Art-making provides patients with a structured yet expressive outlet where internal experiences can be externalized, reorganized, and meaningfully interpreted. This process may restore a sense of agency and identity, particularly in individuals whose illness has disrupted their perception of control.

On the neurocognitive level, this paper examines how creating art may stimulate neural pathways associated with cognition, brain connectivity, and identity. By synthesizing these perspectives, this paper frames art therapy not merely as an emotional aid, but as a comprehensive multisystem approach that yields measurable psychological and neurological benefits.

2.2 Rationale

Individuals undergoing long-term medical treatment or recovery often experience significant emotional and cognitive disruptions, including anxiety, fear, and emotional distress, alongside impairments in memory, attention, and identity coherence (McDougall et al.). A dual analytical framework integrating psychological and neurocognitive perspectives allows for a more comprehensive examination of the recovery process. Specifically, it enables the concurrent analysis of emotional regulation, cognitive engagement, neural activation, and identity reconstruction across a range of illnesses and injuries. This integrative approach moves beyond a narrow focus on symptom reduction. It explores the underlying mechanisms of recovery, identifies patterns across patient populations, and clarifies how creative engagement bridges the gap between emotional and cognitive rehabilitation. Therefore, this combined psychological-neurocognitive analytical framework provides the most appropriate lens for examining the interconnected processes that shape emotional, cognitive, and functional recovery.

3. Background

3.1 Art Therapy as a Therapeutic Intervention

Art therapy is a form of psychotherapy that uses visual art-making, such as drawing, painting, sculpture, and mixed media, within a therapeutic relationship to support emotional, psychological, and cognitive functioning (Joschko et al.). Trained art therapists typically facilitate these interventions, guiding patients through creative processes designed to promote emotional expression, reflection, and adaptive coping. In clinical practice, art therapy is

delivered across a wide range of treatment environments, including hospitals, rehabilitation centers, schools, and community-based programs (Iguina and Kashan).

Empirical investigations of active visual art therapy have been conducted across diverse age groups, including children, adolescents, adults, and older adults, with heterogeneous diagnoses. Study populations frequently encompass individuals with heterogeneous clinical diagnoses, including psychiatric disorders, neurological conditions, oncological illnesses, and chronic medical disease. Reported outcomes across these populations most often target psychological domains such as depression, anxiety, self-esteem, social adjustment, and overall quality of life. In some studies, art therapy has also been evaluated for its potential effects on cognitive functioning, including attention, memory, and executive function (Beadle et al.).

The current evidence for the effectiveness of visual art therapy is derived from a combination of randomized controlled trials, quasi-experimental investigations, pilot trials, and feasibility studies. In addition, systematic reviews and meta-analysis have sought to synthesize these primary investigations across clinical contexts. For example, a systematic review and meta-analysis evaluating approximately fifty clinical trials found that visual art therapy was associated with measurable improvements in several psychological outcomes (Iguina and Kashan). However, the same analysis also highlighted that the magnitude and consistency of observed effects vary substantially across studies. Many outcomes display minimal change, and overall study quality is low (Joschko et al.). Methodological heterogeneity remains a defining characteristic of the art therapy literature, as differences in intervention formats, treatment duration, and outcome assessments make direct comparisons across studies challenging. Consequently, the diversity of study designs and methodological limitations across the field complicate efforts to draw definitive conclusions regarding its effectiveness across clinical indications.

3.2 Psychological Distress and Quality of Life in Oncology

Psychological research in oncology consistently describes high rates of anxiety, depression, emotional distress, and cancer-related stress among patients. These psychological burdens profoundly influence overall quality of life (McPherson et al.). In some cases, art therapy has shown to reduce these symptoms (Jiang et al.). Supporting this, various systematic reviews examining psychological factors in cancer indicate that depression, anxiety, emotional distress, coping strategies, and perceived social support are primary determinants of reported quality of life (Bozcuk et al.). Quality-of-life outcomes in oncology studies are typically measured using standardized multidimensional instruments that assess physical functioning, emotional well-being, social roles, and disease-related symptoms (Abu-Odah et al.). Research on patients undergoing chemotherapy shows that depression can directly worsen quality-of-life outcomes and exert indirect effects through mechanisms such as reduced perceived social support. Long-term treatments, repeated hospitalizations, and uncertainty about prognoses are

often associated with a cumulative emotional burden and cognitive difficulties commonly described as “chemo brain” (Abu-Odah et al.).

3.3 Cognitive Impairment and Identity Disruption in Traumatic Brain Injury

Traumatic brain injury (TBI) is associated with a wide range of emotional and mood-related outcomes. These outcomes commonly include depression, anxiety, irritability, and broader psychological distress. Recent systematic reviews examining self-identity after TBI report that many individuals experience negative changes in self-concept and altered perceptions of their personal identity, shifts that are often linked to heightened emotional distress (Beadle et al.).

Cognitively, TBI can lead to impairments in attention, memory, executive function, and information-processing speed, which may interfere with many aspects of daily and social functioning (Donders and Kim). Moreover, longitudinal neuroimaging studies indicate that TBI can alter connectivity between brain regions involved in autobiographical memory and emotional salience, potentially contributing to fragmented memory processing and altered emotional recall (Beadle et al.). This alteration of sense of self is often described as post-traumatic identity disturbance.

Neuropsychological recovery is typically assessed using standardized cognitive testing, symptom inventories, mood scales, and functional outcome measures, often within longitudinal designs tracking changes over time. In addition to the cognitive scores, outcomes such as social functioning, community reintegration, and role recovery are increasingly recognized as crucial indicators of adjustment.

4. Methods and Materials

4.1 Analysis Framework

This paper adopts an integrative approach that synthesizes findings across psychological and neurocognitive domains to further clarify how different forms of visual art therapy affect mental, cognitive, and functional outcomes. Particular emphasis is placed on its impact on psychological and neurocognitive outcomes in clinical populations, especially individuals experiencing general psychological distress, cancer patients, and individuals with TBI. Examining these populations allows for comparison across clinical contexts in which art therapy may serve different psychological and rehabilitative functions. The analysis is integrative rather than experimental, aiming to compare and interpret findings across diverse research designs to identify consistent patterns, possible mechanisms, and gaps in the existing evidence. This paper analyzes existing review literature and relevant empirical studies on art-based interventions to better understand how visual art therapy influences these different patient groups across various clinical circumstances.

4.2 Sources of Evidence and Selection

Systematic reviews and meta-analyses that synthesize findings across multiple primary studies to estimate overall effect sizes will be prioritized. In particular, this analysis targets systematic reviews and meta-analyses that examine art therapy or other creative art interventions and estimate their overall effects on psychological distress, quality of life, and cognitive outcomes in medical populations. Relevant experimental and quasi-experimental studies are also considered, including randomized controlled trials, controlled clinical trials, and pre-post interventions designs that evaluate structured visual art therapy programs and measure changes in anxiety, mood, coping, memory, or other psychological and cognitive outcomes. Lastly, this paper focuses on clinical and rehabilitation research conducted in hospitals, oncology, pediatrics and rehabilitation settings, including work with cancer patients, children undergoing treatments, or individuals with TBI.

4.3 Analytical and Evaluation Strategy

First, information about the study sample will be extracted, including participant populations such as children with cancer, women with cancer, or TBI patients, while also considering the clinical context, intervention format, and outcome measures used in each study. Outcomes will then be classified into two primary domains: psychological or neurocognitive. Psychological outcomes include anxiety reduction, improved emotional regulation, enhanced coping, and quality of life. Neurocognitive outcomes include attention, memory, executive functioning, and broader indicators of functional recovery. The direction and magnitudes of change within and across populations will be compared. More specifically, the analysis will examine whether visual art therapy consistently reduces psychological distress, strengthens emotional regulation and coping, or supports cognitive stability.

5.1 Discussion

5.1 Framework of Analysis

Based on the literature reviewed in Sections 4.1-4.3, effective non-pharmacological therapies for patients with cancer and traumatic brain injury operate through three core pillars: emotional regulation, identity reconstruction, and neurocognitive engagement.. These domains form a practical framework through which the effectiveness of art therapy can be evaluated. Utilizing this model, art therapy can support psychological stabilization, self-reconstruction, and cognitive engagement within clinical settings.

5.2 Art Therapy as Emotion Regulation

Art therapy may support recovery through interconnected physiological and psychological mechanisms that help patients externalize and regulate difficult emotions (Joschko et al.). The creative process enables nonverbal emotional expression, which may reduce distress and foster adaptive coping. Research reports positive outcomes for patients experiencing mental health challenges, including anxiety and depression, when engaging in art therapy (Joschko et

al.). Beyond offering a nonverbal outlet, art therapy creates a symbolically rich space to externalize emotions that are typically difficult to articulate. This focus is particularly relevant for cancer and TBI patients adapting to life during or after intensive treatments. Additionally, patients experiencing instability can regain a sense of control by utilizing art therapy as a structured tool to restore calm and personal agency. Clinically, increased control reduces distress, strengthens coping mechanisms, and enhances patients' capacity to manage their recovery (Yeom and Lee).

For pediatric and adult female cancer patients, medical treatments often entail prolonged hospitalization, physical discomfort, and a loss of personal control, all of which can compromise emotional regulation (Bozcuk et al.). Art therapy provides a safe and expressive outlet for patients to externalize fears and reconstruct a sense of agency, directly contributing to a reduction in anxiety. Art therapy may activate the parasympathetic nervous system, reducing physiological indicators of anxiety (Zhou et al.). This effect is clinically important, as lowering physiological arousal makes emotional distress more manageable and helps patients maintain a calmer state during treatments. Recent meta-analyses demonstrate that art therapy significantly reduces anxiety in both adult and adolescent cancer patients (Bozcuk et al.). Similarly, studies have reported significant reductions in depressive symptoms among cancer patients utilizing art therapy (Bozcuk et al.). These findings suggest that art therapy may improve self-awareness while enhancing emotion regulation (Zhou et al.), promoting better management of depressive symptoms. This is highly clinically meaningful: reducing anxiety and depressive symptoms improves overall quality of life, increases treatment participation, and supports psychological healing.

For children specifically, the nonverbal and playful nature of art-making may facilitate stress reduction without relying on advanced verbal processing. For women, art therapy supports emotional resilience by reaffirming creativity and identity while lowering stress during physically and psychologically taxing treatments (Iguina and Kashan). Moreover, emotional distress is currently recognized as the sixth vital sign in cancer care, emphasizing the critical importance of assessing well-being alongside traditional physiological measures (Huang and Yin). This clinical need can be addressed non-pharmacologically through art therapy, which provides a safe space for patients to process fear, regulate distress, and rediscover aspects of identity.

5.3 Art Therapy as Identity Reconstruction

Art therapy may also support identity reconstruction by allowing individuals to experiment with new self-narratives and integrate pre-illness and post-illness experiences into their recovery processes. This is critical for cancer and TBI patients, who frequently experience profound identity loss, changes in self-concept, and emotional disruption related to their condition (Beadle et al.). Clinically, a stronger sense of identity may improve coping, reduce psychological fragmentation, and support adjustment during treatment and recovery (Beadle et al.). For TBI patients, identity reconstruction is especially significant. Systematic reviews on

self-identity after TBI report that many individuals experience negative changes in self-concept with altered perceptions of who they are, often accompanied by emotional distress (Beadle et al.). This post-traumatic identity disturbance may leave a long-term effect on the cognitive and somatic domains of an individual's personality (Lanius et al.). Art therapy can help patients process these disruptions in a more accessible way, in comparison to direct verbal explanation alone (Joschko et al.).

For cancer patients, art therapy supports emotional resilience by reaffirming creativity and identity while reducing stress during physically and psychologically taxing treatment (Jiang et al.). In this way, art therapy functions not only as a form of emotional release, but also as a means for patients to reclaim agency and restore a coherent sense of self (Joschko et al.). Restoration of this sense of self is clinically meaningful, as identity continuity supports adaptation, motivation, and long-term recovery.

5.4 Art therapy as Neurocognitive Engagement

Art therapy facilitates both psychological and neurocognitive engagement. It activates attention and motor planning while enabling the means to process identity disruptions caused by neurological shifts (Zhou et al.). A systematic review and meta-analysis suggest that art therapy stimulates the brain's right and left-hemispheres, which can allow the patients to explore complex emotion or non-verbal expression (Zhou et al.). The right hemisphere is often associated with visuospatial and emotional expression, while the left hemisphere contributes to symbolic organization and planning (Zhou et al.). This cross-hemispheric engagement fosters coordinated processing between emotional and cognitive domains (Zhou et al.). This is crucial clinically because stronger coordination between emotional expression and cognitive organization may help patients recover more flexible thinking and better self-regulation during recovery from illness. Recent research has concluded that creative thinking is a two-step process involving idea generation and idea evaluation (Beaty et al.). Idea generation correlates with diffused attention, while idea evaluation usually involves focusing attention and cognitive control (Beaty et al.). This strongly supports the idea that creative cognition utilizes support from both the default and control networks within the brain (Beaty et al.). As shown in Figure 1 (A), whole-brain functional connectivity demonstrates the widespread neural activation required for creative, divergent processing. Additionally, as shown in Figure 1 (B), seed-based connectivity illustrates the functional relationship between the default mode network (PCC), the salience network (insula), and the executive control network (DLPFC). Lastly, Figure 1 (C) displays dynamic posterior cingulate connectivity over time, directly illustrating how connections among brain regions shift across different moments of creative cognition. In the context of art therapy, these patterns may help explain how creative engagement recruits neural systems involved in attention, emotional processing, self-reflection, and cognitive organization. Together, these findings suggest that utilizing art and creativity as a therapeutic mechanism may engage neural networks involved in attention, emotion, memory, and cognitive control.

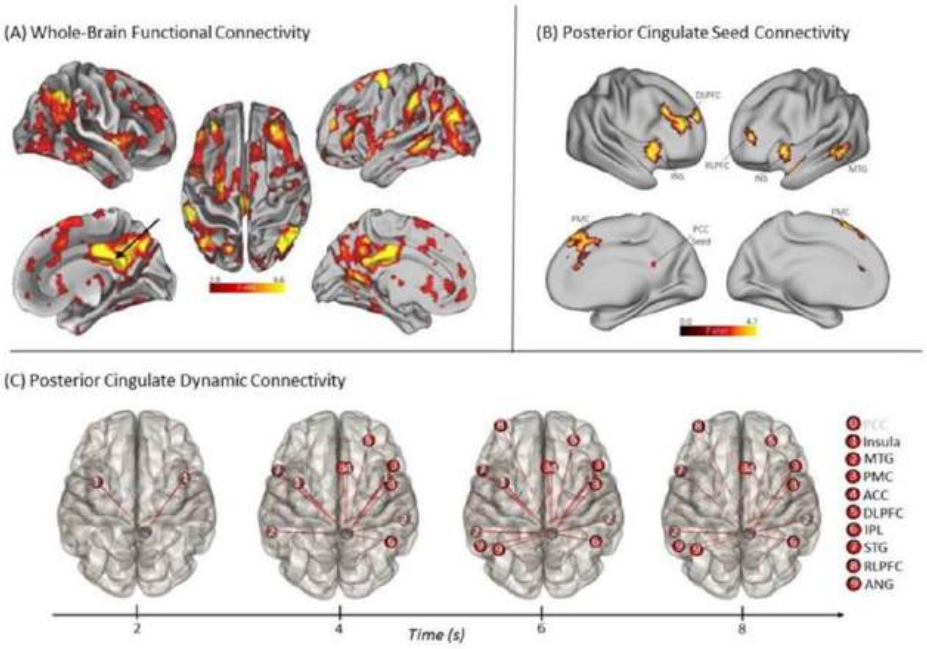


Fig. 1: The Neural Dynamics of Divergent Thinking. From "Creative Cognition and Brain Network Dynamics," by Roger Beaty et al.

Combined Art-Based Interventions (CAI) provide general benefits to cognitive function, language, and memory (Yan et al.). Yan and colleagues also suggest that aging brains retain plasticity in sensory and motor cortices as well as in higher-level cognitive domains, including language, memory, and executive function. Groups participating in art therapies demonstrated significant increases in functional connectivity (FC) within areas such as the temporal and cerebellar anterior lobes and perirhinal and visual cortices (Yan et al.). Together, these changes in connectivity and cognitive performance suggest that art therapy may support neuroplasticity and recovery beyond emotional expression alone. The process of creating art can promote neuroplasticity by repeatedly engaging pathways involved in perception, coordination, attention, and emotional integration (Demarin et al.). This form of cognitive and emotional modulation strengthens connectivity between the limbic system and prefrontal regions, supporting more flexible emotional responses (Sun et al.). As shown in Figure 2, the brain utilizes a specific control center to manage and organize complex emotions. The anatomical mapping (Fig. 2 (A)) traces the structural pathways originating in the right dorsomedial prefrontal cortex (dmPFC) and extending to the peripheral regions. Complementing this, the effective connectivity matrix (Fig. 2 (B)) quantifies the precise strength and directional flow (j → i) of this neural information, proving that the right dmPFC acts as the primary top-down regulator. When a patient engages in art therapy, they must navigate and resolve highly ambiguous emotions. This cognitive process is driven by the significant directional pathways shown in Fig. 2. Instead of allowing bottom-up emotional distress to overwhelm the patient, this prefrontal-parietal network steps into structure,

analyzes, and calms those feelings through this creative expression. By repeatedly demanding this cognitive engagement, art therapy leverages neuroplasticity to strengthen these top-down pathways, facilitating neurocognitive recovery.

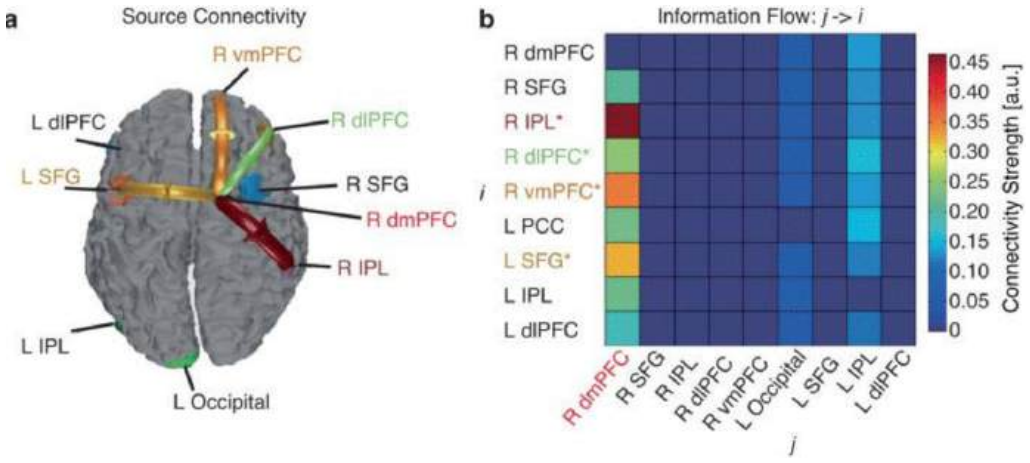


Fig. 2: Neural Connectivity Network Underlying the Processing of Emotional Ambiguity. From “Functional Connectivity Between the Amygdala and Prefrontal Cortex Underlies Processing of Emotional Ambiguity," by Sai Sun et al.

Neuroplasticity allows the brain to forge new connections and activate novel pathways (Demarin et al.). Sustained art-making encourages this plasticity by repeatedly activating attention networks, sensory-motor pathways, and associative cortices, reinforcing adaptive connections (Demarin et al.). By activating these attention, planning, and emotional processing networks, art therapy stimulates cognitive and emotional processing through creative engagement in ways that complement traditional neurorehabilitation and psychosocial interventions. Emotional expression enhances attention and motivation, while cognitive engagement reinforces self-efficacy and coherence of identity (Demarin et al.).

This neurocognitive interpretation is further supported by research showing that visual art production may alter functional connectivity within brain networks involved in self-reflection, attention, and cognitive control. As shown in Figure 3, Bolwerk et al. compared brain connectivity before a 10-week visual art intervention at T0 and after the intervention (T1). The T1 > T0 comparison highlights regions where connectivity increased following the intervention. Participants who actively produced visual art showed broader increases in connectivity between the posterior cingulate/precuneus and frontal-parietal brain regions than participants who only evaluated artwork. These findings suggest that active art-making may influence neural systems involved in self-reflection, emotional regulation, attention, and resilience.

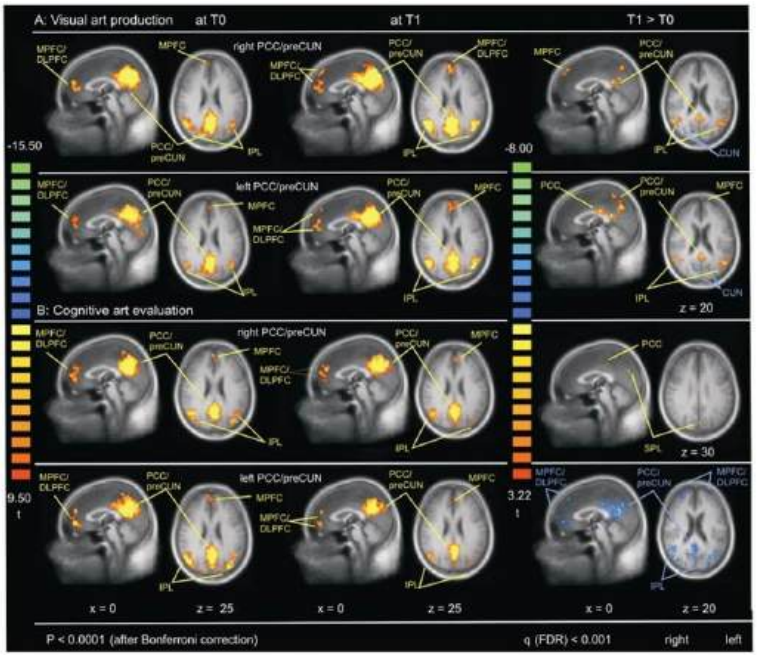


Fig. 3: Changes in default mode network connectivity after visual art production. From "How Art Changes Your Brain: Differential Effects of Visual Art Production and Cognitive Art Evaluation on Functional Brain Connectivity," by Anne Bolwerk et al.

Overall, art therapy operates through interconnected mechanisms that support emotional regulation, identity reconstruction, and neurocognitive engagement. These three mechanisms provide a clear framework for understanding how art therapy may contribute to healing in clinical settings. Rather than acting as a single isolated treatment effect, art therapy appears to support recovery by addressing distress, self-concept, and cognitive functioning simultaneously.

Remaining Gaps

A key limitation in the current literature is the scarcity of large-scale, well-controlled clinical trials, particularly for specific subgroups such as pediatric oncology patients, women with cancer, and individuals navigating TBI (Zhou et al.). There is also a lack of standardized intervention models with wide variation in session length, frequency, materials, therapeutic orientation, and outcome measures (Joschko et al.). This makes it difficult to compare studies of synthesizing reliable estimates. Moreover, long-term outcome data is limited, as most studies assess short-term changes in mood or distress, with few examining sustained effects on identity, social functioning, or neurocognitive recovery within months or years after treatments (Joschko et al.). Investigating art-based interventions within clinical populations, such as children and women undergoing cancer treatments, and with individuals recovering from TBI, can be crucial for understanding the multidimensional benefits of art therapy in the years to come. These benefits show great correlation with mental health challenges, as "art therapy can positively improve psychological outcomes, particularly anxiety and depression" (Zhou et al.). However,

limited study availability and methodological heterogeneity weaken the current evidence base (Zhou et al.).

Future Directions

Future research should prioritize longitudinal clinical trials where they can follow patients across key phases of illness and recovery, examining whether art therapy produces durable improvements in anxiety, depression, coping, and quality of life (Beadle et al.). Furthermore, neuroimaging studies could clarify how creative engagement influences brain networks related to attention and emotional regulation. A lens focused on art therapy-related creative engagement can lead to possible advancements in patient treatment processes. Through cross-population comparisons, research can identify more tailored intervention protocols, specifically determining for whom and under what conditions is art therapy most effective. Additionally, future studies could explore the cost-effectiveness of art therapy interventions or the integration of digital or virtual art therapy platforms. Further research with art therapy interventions has extensive possibilities and outcomes for patients if pursued.

Additional Considerations

The integration of art therapy into hospitals and rehabilitation centers–such as bedside sessions in oncology wards or structured group programs in neurorehabilitation settings–highlights its potential to make psychosocial support more accessible to patients in clinical environments (Jiang et al.). These findings carry significant implications for healthcare policy, suggesting that creative therapies should be formally incorporated into funding priorities, staffing models, and multidisciplinary care approaches, particularly in settings where access to traditional psychological services is limited. Additionally, the role of creative therapies extends beyond treatment into prevention. Embedding art therapy-based programs into primary care settings could build emotional resilience before serious illness or injury occurs. In this way, art therapy serves as a bridge between clinical intervention and broader public health promotion.

Works Cited

Beadle, Elizabeth Jane, et al. “The Impact of Traumatic Brain Injury on Self-Identity.” Journal of Head Trauma Rehabilitation, vol. 31, no. 2, Mar. 2016, pp. E12–E25, https://doi.org/10.1097/htr.0000000000000158.

Beaty, Roger E., et al. “Creative Cognition and Brain Network Dynamics.” Trends in Cognitive Sciences, vol. 20, no. 2, Feb. 2016, pp. 87–95, https://doi.org/10.1016/j.tics.2015.10.004.

Bolwerk, Anne, et al. “How Art Changes Your Brain: Differential Effects of Visual Art Production and Cognitive Art Evaluation on Functional Brain Connectivity.” Plos One, vol. 9, no. 7, 1 July 2014, https://doi.org/10.1371/journal.pone.0101035.

Bozcuk, H., et al. “A Comparative Study of Art Therapy in Cancer Patients Receiving Chemotherapy and Improvement in Quality of Life by Watercolor Painting.” Complementary Therapies in Medicine, vol. 30, Feb. 2017, pp. 67–72, https://doi.org/10.1016/j.ctim.2016.11.006.

Demarin, Vida, et al. “Arts, Brain and Cognition.” Psychiatria Danubina, vol. 28, no. 4, 1 Dec. 2016, pp. 343–348, pubmed.ncbi.nlm.nih.gov/27855424/.

Donders, Jacobus, and Eunice Kim. “Effect of Cognitive Reserve on Children with Traumatic Brain Injury.” Journal of the International Neuropsychological Society, vol. 25, no. 04, Apr. 2019, pp. 355–361, https://doi.org/10.1017/s1355617719000109.

Hammoda Abu-Odah, et al. “Effectiveness of Creative Arts Therapy for Adult Patients with Cancer: A Systematic Review and Meta-Analysis.” Supportive Care in Cancer, vol. 32, no. 7, 14 June 2024, https://doi.org/10.1007/s00520-024-08582-4.

Huang, Tangsen, and Xiangdong Yin. “Exploring the Benefits of Art Therapy for Anxiety and Depression in Women with Breast Cancer: A Meta-Analytical Systematic Review.” Journal of Psychiatric Research, May 2025, https://doi.org/10.1016/j.jpsychires.2025.05.027.

Iguina, Michele M., and Sanaz Kashan. “Art Therapy.” PubMed, StatPearls Publishing, 2021, pubmed.ncbi.nlm.nih.gov/31747178/.

Jiang, Xiao‐Han, et al. “Effects of Art Therapy in Cancer Care: A Systematic Review and Meta‐Analysis.” European Journal of Cancer Care, vol. 29, no. 5, 15 June 2020, https://doi.org/10.1111/ecc.13277.

Joschko, Ronja , et al. “Active Visual Art Therapy and Health Outcomes.” JAMA Network Open, vol. 7, no. 9, 12 Sept. 2024, pp. e2428709–e2428709, jamanetwork.com/journals/jamanetworkopen/fullarticle/2823638?resultClick=1#google_vignette, https://doi.org/10.1001/jamanetworkopen.2024.28709.

Lanius, Ruth A., et al. “The Sense of Self in the Aftermath of Trauma: Lessons from the Default Mode Network in Posttraumatic Stress Disorder.” European Journal of Psychotraumatology, vol. 11, no. 1, 23 Oct. 2020, p. 1807703, www.ncbi.nlm.nih.gov/pmc/articles/PMC7594748/, https://doi.org/10.1080/20008198.2020.1807703.

Manning, Joseph C, et al. “Long-Term Psychosocial Impact Reported by Childhood Critical Illness Survivors: A Systematic Review.” Nursing in Critical Care, vol. 19, no. 3, 23 Oct. 2013, pp. 145–156, https://doi.org/10.1111/nicc.12049.

Martinelli, Valentina, et al. “Active Visual Art Therapy in the General Hospital: Facts and Challenges from an Ethical Perspective.” International Journal of Environmental Research and Public Health, vol. 22, no. 2, 19 Feb. 2025, pp. 316–316, https://doi.org/10.3390/ijerph22020316.

McDougall, Graham J., et al. “Memory and Cancer: A Review of the Literature.” Archives of Psychiatric Nursing, vol. 28, no. 3, 1 June 2014, pp. 180–186, www.ncbi.nlm.nih.gov/pmc/articles/PMC4033831/, https://doi.org/10.1016/j.apnu.2013.12.005.

McPherson, Susan, et al. Quality-of-Life Measurement in Depression Trials: A Consumerist Relic. 28 Jan. 2022, p. 136345932210748-136345932210748, https://doi.org/10.1177/13634593221074887.

Sun, Sai, et al. “Functional Connectivity between the Amygdala and Prefrontal Cortex Underlies Processing of Emotion Ambiguity.” Translational Psychiatry, vol. 13, no. 1, 28 Oct. 2023, pp. 1–14, www.nature.com/articles/s41398-023-02625-w, https://doi.org/10.1038/s41398-023-02625-w.

Yan, Yuan-Jiao, et al. “Impact of Combined Art-Based Intervention on Functional Connectivity of Multiple Brain Networks in Older Adults along the Cognitive Continuum: Result from a Parallel Randomised Controlled Trial.” BMC Psychiatry, vol. 25, no. 1, 31 Mar. 2025, https://doi.org/10.1186/s12888-025-06741-3.

Yeom, Hyun-E, and Jungmin Lee. “Impact of Autonomy Support on the Association between Personal Control, Healthy Behaviors, and Psychological Well-Being among Patients with Hypertension and Cardiovascular Comorbidities.” International Journal of Environmental Research and Public Health, vol. 19, no. 7, 31 Mar. 2022, p. 4132, https://doi.org/10.3390/ijerph19074132.

Zhou, ShiShuang, et al. “The Effects of Art Therapy on Quality of Life and Psychosomatic Symptoms in Adults with Cancer: A Systematic Review and Meta-Analysis.” BMC Complementary Medicine and Therapies, vol. 23, no. 1, 1 Dec. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10691094/#:~:text=Several%20studies%20have%20indicated%20that, https://doi.org/10.1186/s12906-023-04258-4.

Donation Transparency and Consumer Skepticism in response to Cause-Related Marketing
By Misha Cherukumilli

Abstract

As cause-related marketing (CrM) becomes more common, companies use charitable claims to signal social responsibility. However, consumers' responses to these messages are not uniform. This study examines whether explicitly stating a donation percentage increases consumer trust and purchase intent, and how skepticism and persuasion knowledge moderate this relationship. It was hypothesized that stating "we donate X%" would increase purchase intent and trust only when the donation claim is transparent and verifiable, and would be less effective among consumers with higher persuasion knowledge and skepticism. This hypothesis was tested using an experimental survey with three conditions: no donation claim (control), a vague donation claim, and a specific donation claim stating that 5% of purchases would be donated to the American Red Cross. Participants reported purchase intent, trust, skepticism, and perceived transparency using 7-point Likert scale measures. Results showed that explicit donation claims increased perceived detail and modestly increased average trust and purchase intent compared to vague or absent donation claims. However, responses remained polarized, and many participants still reported low perceived verifiability. These findings indicate that specificity improves clarity, but it does not consistently eliminate skepticism or guarantee consumer trust.

Keywords: CrM, consumer skepticism, donation transparency, persuasion knowledge, advertising credibility, adolescent consumers

Introduction

Cause-related Marketing (CrM) campaigns have emerged as a major source of consumer spending, as demonstrated by the American Express Statue of Liberty campaign in 1983, one of the first CrM campaigns, which generated $1.7M. This campaign was an early demonstration of what would later be termed as CrM, a marketing strategy that links consumer spending to charitable donations. CrM is more specifically defined in 1988 by Varadarajan and Menon in their seminal article as "the process of formulating and implementing marketing activities that are characterized by an offer from the firm to contribute a specified amount to a designated cause when consumers engage in revenue-providing exchanges that satisfy organizational and individual objectives" (p. 60). Despite the early success of CrM, subsequent research suggests that consumer responses to charitable claims are not uniformly positive. Branching off from broad corporate social responsibility (CSR), CrM is a specific form that differs from traditional philanthropy. The company benefits first before the obligation to donate is incurred, which may cause consumers to view the charitable claim as less altruistic and lead CrM to backfire. This can activate consumer skepticism, which is defined as the general tendency to doubt the truthfulness or sincerity of marketing claims. Consumer skepticism is an important variable because it plays a critical role in shaping consumer purchase intention and trust in CrM contexts. Consumers have

developed greater persuasion knowledge as CrM becomes more widely practiced, enabling them to critically evaluate donation claims. This is especially prominent when such claims are vague or difficult to verify.[6] Existing literature cumulatively suggests that the effectiveness of CrM is dependent not only on whether a donation is promised, but also on how credible and transparent the promise is communicated.

Importantly, transparency and verifiability are two distinct concepts. Transparency refers to how clearly donation information is presented, such as stating the percentage donated or naming a beneficiary. Verifiability refers to whether consumers feel they have sufficient information to confirm that the donation will actually occur. Prior research suggests that increasing message detail may improve perceived transparency, but does not necessarily eliminate skepticism if consumers still question whether the donation is real or enforceable. This is a critical distinction because donation claims may appear more detailed without being perceived as fully credible.

Building on this literature, this study examines whether explicitly stating a donation percentage increases purchase intent or instead backfires by increasing consumer skepticism, as well as the moderating roles of perceived transparency, perceived verifiability, and persuasion knowledge on the overall relationship. Due to the growing popularity of CrM campaigns, understanding how consumers respond to explicit donation claims has become increasingly important due to financial stakes. CrM campaigns influence large shares of consumer purchasing decisions and channel a significant amount of consumer spending through donation-linked promotions. Making a small tweak in the donation message framing can be consequential for both firms and the charitable cause to which the donation is given. Prior studies show that CrM can positively influence purchase intentions, but consumers often misunderstand donation structures and evaluate products in inaccurate ways, particularly when donation information is ambiguously presented. Such misinterpretations are relevant because they can distort purchasing behavior, undermine trust in brands, and inflate the perceived social impact of their purchase. Determining whether explicitly stating a donation percentage clarifies information and builds trust or increases consumer skepticism represents an unresolved issue in CrM literature. We hypothesize that stating a donation percentage increases purchase intent and trust only when the claim is transparent and verifiable, and is less effective or backfires among consumers with higher persuasion knowledge and skepticism. To test this hypothesis, the study used three condition experimental design, which features a hypothetical smartwatch brand presented across one of three conditions: a control condition with no donation message, a vague donation condition that referenced charitable giving without stating donation percentage or beneficiary, and a specific donation condition that stated donation percentage and recipient. Participants completed a series of 7-point Likert-scale items which measured purchase intent, brand consideration, trust, message believability, consumer skepticism, and persuasion knowledge. A manipulation check was included to assess if participants noticed the donation message. Demographic information, such as age and gender were also collected. To ensure random

assignments of participants to a group, participants were sent to one of three conditions based on the last digit of their phone number.

Methods

This study aims to analyze consumer response to explicit donation claims in comparison to vague donation claims, and the factors that affect this response, such as transparency, persuasion knowledge, consumer skepticism, and verifiability. This study uses a survey-based experimental design where participants were shown a hypothetical smartwatch brand named Vero Watch. They were given a short description of the watch brand, about its use, and the price of $79. This price was chosen because it falls in the median range of general smartwatch prices currently on the market. They were then prompted to answer a series of 7-point Likert Scale and demographics questions.

Participants were recruited through convenience sampling in the Palo Alto and greater Bay Area during the period of Dec 2025 to Jan 2026 and completed the survey voluntarily. All participants confirmed they were at least 13 years of age before responding. Participants were required to be at least 13 years of age because early adolescents generally possess the cognitive capability to comprehend survey questions, making those around the age of 13 and above appropriate participants for consumer research (Steinburg, 2005; Kuhn, 2006) Responses were anonymous, and no personal information was collected.

Participants were assigned to treatment conditions using a rule-based assignment procedure determined by the last-digit of their phone number, rather than platform random assignment. This method ensured a systematic and unbiased distribution across conditions while allowing participants to be categorized consistently into one of the three experimental groups.

Treatments were various sections of the surveys with different information about donation claims. According to the Internal Revenue Service (IRS), donations are around 1-2% and capped at a maximum of 10% (IRS, 2026). Business donations typically do not exceed 5%. For example, Patagonia donates 1% of its annual sales to environmental organizations (Patagonia, 2026), while PVH Corp. donated 3.8% of its pre-tax profits in a single reported year (Vogue Business, 2019). Therefore, this study used a charitable donation claim of 5% in the explicit donation claim condition to demonstrate a significant contribution. Treatment group 1 was the control with no donation claim (see Figure 1) and assigned to participants with phone numbers ending in 0-3. Treatment group 2 was the vague donation claim with no explicit percentage donated or beneficiary stated. The vague donation claim said “We donate a portion of profits to charity” (see Figure 2) Participants in treatment group 2 had phone numbers ending in 4-6. Treatment group 3 was the explicit donation claim with a specific beneficiary and the percentage donated stated. The explicit donation claim stated “We donate 5% of this purchase to the American Red Cross” (see Figure 3). Participants in treatment group 3 had phone numbers ending in 7-9.



Figure 1. Control Treatment (No Donation Claim)



Figure 2. Vague Donation Claim Treatment



Figure 3. Explicit Donation Claim Treatment

The primary material used in this study was an online Google Forms survey. The survey included:

- An informed survey/eligibility check
- Brand stimuli describing a hypothetical smartwatch product.
- 7-point Likert Scale questions measuring purchase intent, trust, believability, and skepticism
- Questions assessing perception of transparency
- Demographics questions

Participants viewed images and text describing Vero Watch. Each treatment had varying information about Vero Watch's donation claims.

Participants first completed an eligibility question confirming their willingness to participate. They were shown an image of the Vero Watch brand and given a short description and, depending on the treatment group assigned, shown no donation claim, a vague donation claim, or a specific donation claim. After viewing brand information, participants answered Likert-scale questions evaluating their likelihood of purchasing the product, their persuasion knowledge, skepticism, and perceptions about donation transparency. Participants completed demographic questions about age and gender before completing the survey.

Most variables were measured using a 7-point Likert scale (see Table 1 for scale anchors). Anchors ranged from (1) Strongly Disagree to (7) Strongly Agree or (1) Very Unlikely to (7) Very Likely. These variables were used to measure skepticism, consumer persuasion knowledge, and purchase intention. Perceived transparency was measured using questions assessing whether participants believed the donation information provided sufficient detail and clarity. The questions asked using the Likert scale include the following: Some example

questions asked the following: “The donation information provided includes concrete details” and “There is enough information to determine whether the donation occurs.”

Persuasion knowledge was assessed using Likert scale items measuring participants’ tendency to question the truthfulness and intent of the brand’s donation claims. This approach is consistent with the Persuasion Knowledge Model which defines persuasion knowledge as consumers’ ability to recognize and evaluate persuasive attempts, and established skepticism scales that measure doubt toward marketing claims.[2,9] Higher scores indicated greater skepticism and critical evaluation of the donation message.

Measures

Measures were designed to assess consumer responses to donation transparency across multiple constructs, including purchase intent, brand trust, skepticism, persuasion knowledge, and perceived transparency. All measures were assessed using 7-point Likert scales, with higher values indicating greater agreement or stronger endorsement of the construct being measured (see Table 1).

Purchase intent was measured using questions which assessed participants’ likelihood of purchasing the Vero Watch product. Brand trust was measured using items evaluating perceived credibility and reliability

Analysis

This paper uses descriptive statistics by analyzing the mean, standard deviation, and median of the 3 treatment groups: control with no donation, vague donation claim, explicit donation claim. The groups were compared using the descriptive statistics to evaluate any differences in purchase intent, trust, skepticism, perceived transparency, and perceived verifiability. The response distributions were also compared using the percentage of participants selecting low (1-2), neutral (3-5), and high (6-7) on the Likert scale.

Moderation Analysis

Role of Persuasion Knowledge

To examine whether persuasion knowledge moderated responses to donation transparency, participants were divided into two groups based on their responses to persuasion knowledge items. Participants who selected higher agreement values (5-7 on the Likert scale; see Table 1) were categorized as having relatively high persuasion knowledge which indicated greater awareness of persuasive intent and marketing tactics. Participants selecting lower agreement values (1-3) were categorized as having relatively lower persuasion knowledge.

Participants with lower persuasion knowledge showed a clearer increase in trust and purchase intent in response to the specific donation claim compared to the vague and control treatments. In contrast, participants with higher persuasion knowledge did not show a consistent increase in trust or purchase intent in the specific donation condition. Many participants in the

high persuasion knowledge group continued to report low trust and high doubt despite the presence of a specific donation percentage and a named beneficiary.

This pattern suggests that the effect of donation transparency differed depending on persuasion knowledge level. For participants with lower persuasion knowledge, providing specific donation details appeared to improve perceived credibility and increase willingness to consider purchasing a product. However, for participants with higher persuasion knowledge, increased specificity did not uniformly improve trust of the brand. Skepticism was measured using agreement with statements such as "I have doubts about the claims made by this brand."

Results & Discussion

Demographic Overview

Survey responses from 62 participants were collected and analyzed. Participants were randomly assigned to one of three treatments based on the last digit of their phone number: control (no donation claim), vague donation claim, or explicit donation claim stating that 5% of the purchase will go to the American Red Cross. Responses were measured using a 7-point Likert scale.

The sample was not evenly distributed across the demographic groups (Table 1). Out of the 62 participants, 45 identified as female (72.6%) and 17 identified as male (27.4%). Age distributions were also uneven. There were zero under 14 years (0.0%), 23 aged 14-17 years (37.1%), zero aged 18-24 (0.0%), three aged 25-34 years (4.8%), 31 participants aged 35 and older , and five participants (8.1%) selected "Prefer not to say".

Characteristics	Category	Frequency	Percentage
Sex	Female	45	72.6%
	Male	17	27.4%
Age	Under 14 years	0	0.0%
	14-17 years	23	37.1%
	18-24 years	0	0.0%
	25-34 years	3	4.8%
	35+ years	31	50.0%
	Prefer not to say	5	8.1%

Table 1: Participant demographics

Across all treatments, respondents had varying levels of purchase intent and likelihood to buy Vero Watch (hypothetical brand). This indicates varying impressions of the Vero Watch brand as a whole for various respondents.

Across the three experimental conditions, purchase intent ratings exhibited substantial variability rather than clustering around a single central tendency. In both the control and vague donation conditions, responses ranged widely across the 7-point Likert scale. This indicates that participants had varying impressions of the Vero Watch hypothetical brand.

Purchase Intent and Consideration

Purchase intent differed across the three experimental conditions (Figure 4). In the control group, average purchase intent was M = 2.86 (SD = 1.98, median = 2), and a majority of participants selected a very low rating on the 7-point likert scale: 57.1% chose 1-2 (chose close to "Very Unlikely" on scale), while 9.5% chose 6-7 (chose close to "Very Likely" on scale). The vague donation claim condition had an average purchase intent increased to M = 3.53 (SD = 1.70, median = 3). However, responses still polarized across the full 7-point Likert scale: 23.5% selected 1-2, and 17.6% chose 6-7. The specific donation claim condition stated that "5% of purchases to the American Red Cross" had the highest purchase intent on average, with a mean of 3.96 (SD = 2.16, median = 4). However, this increase did not reflect a uniform shift toward higher ratings. It is important to note that the distribution was still polarized rather than uniformly positive. 29.3% of respondents selected 1-2 (strongly disagree to disagree), while 25% selected 6-7 (agree to strongly agree), indicating a split between highly interested and highly uninterested participants even under transparent and specific claims.



Figure 4. Mean Purchase Intent by Donation Transparency

Skepticism and Trust-related measures

Table 2 summarizes descriptive statistics for purchase intent, skepticism, and trust across the three experimental conditions. Trust showed smaller shifts than purchase intent. In the control group, trust had a mean of 3.38 on the 7-point Likert scale (SD = 1.88, median = 3), 33.3% selecting 1-2 (chose close to "Strongly Disagree" on scale), on the scale and 19.0% selecting 6-7 (chose close to "Strongly Agree" on scale). In the vague donation claim group, the average score for trust was similar, with a mean of 3.41 (SD = 1.80, median = 3). 29.4% of respondents selected 1-2, and 17.6% selected 6-7. Agreement with the question that stated "I have doubts about the claims made" was highest in this condition (M = 4.41, SD = 1.80, median = 5). 35.3% of participants selected 6-7 on the Likert scale, which suggests that vague donation language coincided with more profound expressed doubt from respondents. In the specific donation claim group, trust increased modestly (M = 4.17, SD = 1.55, median = 4). Low trust ratings were less common in this group, with 8.3% selecting 1-2. However, the trust distribution did not become overwhelmingly high and remained polarized (12.5% selected 6-7 on the scale). These results indicate that specifying a beneficiary and percentage does not automatically produce strong trust for most respondents.

Condition	Population (N)	Likelihood to Buy (7-point Likert Mean)	Trust in Vero Watch (7-point Likert Mean)	Doubts About Claims (7-point Likert Mean)
Control (No Donation)	21	2.86	3.38	3.90
Vague Donation Claim	17	3.53	3.41	4.41
Explicit Donation Claim	24	3.96	4.17	3.83

Table 2. Descriptive Statistics by Donation Transparency Condition

Perceived Transparency and Manipulation Check

The manipulation check question was included at the end of the vague donation group's and specific donation group's survey. The check indicated that most participants recognized the donation claim: 79.2% of the participants in the specific donation claim condition answered "Yes" to whether Vero Watch mentioned donation money (19 out of 24 participants). 82.4% of participants in the vague donation claim condition answered "Yes" to the same manipulation check.

Perceived transparency of the Vero Watch hypothetical brand clearly separated the two donation conditions (vague vs. specific). Compared to the vague condition, the specific condition was rated as providing more "concrete details". The specific donation claim condition had a mean rating of 4.46, while the vague donation claim condition had a mean rating of 2.29 for the

question that prompted respondents, “The donation information provided by *Vero Watch* includes concrete details”. The specific donation claim condition also seemed to place a stronger emphasis on the charitable cause, based on the participating mean rating of 4.21 compared to the mean rating for the vague condition of 2.06.

Despite this, participants did not uniformly feel they had enough information to verify the donation claim, even in the specific condition. On the survey question stating “There is enough information to determine whether the donation occurs,” the specific condition averaged a mean of 3.17 (median = 2.5), with 50.0% of respondents selecting 1-2 on the 7-point Likert scale. Ratings for the survey question stating “Vero Watch explains what happened to the donation” were also mixed in the specific condition, with a mean rating of 3.75 and substantial responses on both ends of the scale (37.5% selecting 1-2 and 33.3% selecting 6-7). This suggests that stating a percentage and beneficiary improved perceived concreteness and clarity of the donation, but many participants do not view the claim as fully verifiable based on the survey results.

This interpretation was also consistent with prior research which found that consumers who perceive persuasive intent may respond with increased resistance and reduced trust.[10] In this study, participants in the specific donation condition rated the claim as more detailed but still reported relatively low agreement with statements indicating that the donation could be verified. This suggests that while specificity improved perceived clarity, it did not fully overcome credibility concerns among more skeptical patterns.

Demographic patterns further help explain the polarized response distribution. As illustrated by Table 1, the sample was mainly concentrated in two age groups: adolescents aged 14-17 years (37.1%) and participants aged 35 years and older (50.0%), while much fewer respondents were aged 25-34 years (4.8%), and none were aged under 14 years or 18-24 years. Prior research shows that adolescents develop persuasion knowledge and become increasingly sensitive to persuasive intent as they gain experience with marketing (Boush, Friestad, & Rose). As a result, some adolescent participants may have interpreted the explicit donation claims as a persuasive tactic rather than purely altruistic behavior, contributing to skepticism despite increased transparency.

These results provide preliminary evidence that persuasion knowledge influenced how donation transparency was interpreted, supporting the hypothesis that consumer awareness of the persuasive intent moderated responses to CrM claims.

The purpose of this study was to examine whether increasing donation transparency in CrM improves consumer trust and purchase intent, or whether it instead activates skepticism among consumers. It was hypothesized that explicitly stating a specific donation percentage and beneficiary would increase both trust and purchase intent, but only when the claim was perceived as credible and verifiable by the consumer. This study also hypothesized that this effect would be weaker among consumers with higher persuasion knowledge. The results partially supported this hypothesis.

As shown in Table 2 and Figure 1, the explicit donation claim treatment produced the highest average purchase intent (M = 3.96) and trust (M = 4.17) compared to the vague donation claim and control treatments. However, these increases were modest and did not produce uniformly positive responses across all participants, Even in the explicit donation condition, a substantial number of participants reported very low purchase intent and trust ratings, indicating the transparency did not consistently increase positive customer evaluations.

This study examined whether increasing donation transparency in CrM increases consumer trust and purchase intent, or whether it instead leads to greater skepticism. It was hypothesized that explicitly stating a donation percentage and beneficiary would increase trust and purchase intent only when the claim was perceived as verifiable, and that such claims could be less effective among consumers with higher persuasion knowledge.

The results partially supported this hypothesis. Explicit donation claims had the highest mean level of purchase intent and trust on the 7-point Likert scale measure compared to vague donation claims and no donation claims. However, these increases were modest and inconsistent across participants. Rather than shifting responses uniformly in a positive direction, the explicit donation condition produced polarized outcomes, with some participants responding positively and others remaining highly skeptical. This suggests that increased transparency was not universally persuasive across all respondents.

In contrast, vague donation language seemed to constantly reduce trust among respondents/ Participants given that the vague donation claim condition expressed higher levels of doubt, indicating that ambiguous claims may heighten skepticism by leaving brand motives unclear. This finding also aligned with prior research, which showed that when charitable claims lack specificity, consumers are more likely to attribute this behavior to selfish motives that only benefit the company rather than genuine concern and willingness to help.[3]

Taken altogether, the findings suggested that donation transparency functions as a conditional cue rather than a universally positive signal. While specificity increased trust for some consumers, it simultaneously activated skepticism among others, particularly when claims were not perceived as verifiable.

Limitations

This study has several limitations. Surveys were not randomly distributed, and convenience sampling is present. The survey relies on recruiting participants via personal networks rather than simple random sampling (SRS). As a result, the sample is not representative of the general population, and results cannot be generalized beyond the population that decided to participate. Despite this limitation, this study acts as an informative exploratory analysis of how consumers interpret charitable donations in CrM contexts, and the observed patterns provide insight into the broader mechanisms that shape skepticism and trust. The survey was primarily completed by individuals living in California and may not reflect the perception held by consumers in other regions. The demographic composition of the sample was uneven. Respondents were concentrated in age groups 14-17 and 35+, while other age groups were

minimally represented or not represented. More respondents identified as female compared to male. Consumer trust, persuasion knowledge, and persuasion knowledge are all affected by age, gender, and general demographics, so this imbalance may affect findings and restrict generalizing results to underrepresented age groups within the study.

The survey uses a hypothetical smartwatch brand rather than an existing brand. This allowed for greater experimental control, but responses may not accurately represent real-world consumer behavior. If an existing brand was used, factors such as brand familiarity, prior experience, and external information about the brand may have affected responses.

This study compared several variations of donation transparency across survey groups; varying levels of participant attention could have affected responses. Although a manipulation check was included, varying levels of attention can still reflect responses. Participants were assigned treatments on the last digit of their phone number, which may not represent perfect distribution across all treatments. There is a possibility respondents could not accurately recall the last digit of their phone number or entered an incorrect digit, introducing potential misclassification. Experimental groups may differ in size or composition, and could influence comparison across the donation transparency conditions.

Finally, a test like the Mann-Whitney Test was not completed to establish a statistical difference across means. This study was conducted independently by a high school student and statistical analyses were deliberately limited to descriptive methods such as means, standard deviations, medians, and response distributions. These were appropriate for the exploratory scope of this project. Although this limitation exists, the study still provides preliminary evidence around consumer responses to donation transparency and generally offers a foundation for future research using larger samples and inferential statistical analyses.

Conclusion

Differences in how consumers responded to donation claims across experimental conditions highlighted the complex roles of transparency in CrM. Using a survey-based experimental design, the study compared consumer responses to a no donation claim, a vague donation claim, and an explicit donation claim stating both a percentage and a beneficiary.

The findings indicated that the donation transparency had clear but limited effects. Vague donation claims were associated with heightened skepticism, indicating that a lack of specific information may intensify doubt rather than mitigate it. Explicit donation claims improved perceived clarity and modestly increased trust and purchase intent, but they did not eliminate doubt. Even when donation details were provided, many participants remained uncertain about whether the donation would actually occur.

The results have highlighted a key distinction between clarity and credibility in CrM. While transparency improved consumers' understanding of the actual donation message itself, it did not consistently establish trust. This suggested that effective CrM may require mechanisms that go beyond disclosure, such as verifiable evidence of donation fulfillment, confirmation by the nonprofit beneficiary, or visible accountability, such as real-time donation trackers.

Several broader questions emerged through this study. Future research could examine whether adding verification cues–such as confirmation, visible trackers, etc–reduces skepticism more effectively than transparency alone, particularly among adolescents with high persuasion knowledge. Additional studies could also investigate how these dynamics differ across age groups or when real brands with established reputations are used, instead of the hypothetical brand used in this study: Vero Watch. Experimental designs that independently manipulate transparency and verifiability may help clarify how consumers evaluate informational detail to judge the credibility of charitable marketing claims.

Overall, this study contributed to CrM literature by demonstrating that increased transparency alone did not reliably produce consumer trust. Instead, transparency may activate persuasion knowledge and critical evaluation, particularly among skeptical consumers, resulting in polarized responses instead of uniform increases in purchase intent or trust. This finding helps explain why some CrM campaigns fail to produce the intended positive effects, even when donation claims are clearly communicated.

These findings also have important practice implications for firms using CrM. Simply stating a donation percentage may not be sufficient to build consumer trust. Firms may need to provide additional verification mechanisms, such as third-party nonprofit communication, publicly accessible donation records, or follow-up communication showing donation fulfillment to ensure consumers perceive donation claims as credible rather than purely promotional.

Future research could further investigate how verification mechanisms influence consumer trust, whether these effects differ across age groups, and how persuasion knowledge develops across different demographic populations. Understanding these mechanisms will be important for designing CrM campaigns that are effective and credible.

By distinguishing between transparency and perceived credibility, this study provides new insight into how consumers interpret the donation-based marketing claims and offers guidance for improving the effectiveness of CrM strategies.

Acknowledgements

Dr. Jing Shi's guidance through this process was invaluable. Her thoughtful feedback, one-on-one mentorship, and encouragement strengthened the quality and organization of my work. Additionally her input on the analysis of the results section and overall structure of the paper substantially strengthened the clarity and rigor of the final work. I additionally thank Kieran Tait for his careful review and revision feedback, which significantly improved the precision, organization and academic quality of this project. I am also grateful to Daryn Howland for her prolonged support and mentorship throughout the research process. She provided helpful direction, answered methodological and writing questions, and helped keep the project on track through regular check-ins and accountability. I would like to thank Dr. Elgin for his detailed guidance and thoughtful feedback throughout the process of writing this research project. Finally, I appreciate all survey participants whose responses made this study possible.

Works Cited

Anuar, Marhana Mohamed, and Osman Mohamad. "Examining the Effects of Cause-Proximity and Gender on Consumers' Response to Cause-Related Marketing: Evidence from Malaysia." International Journal of Marketing Studies, vol. 3, no. 3, 2011, pp. 174–181. https://doi.org/10.5539/ijms.v3n3p174

Arnett, George. "How Charitable Are Fashion's Biggest Companies?" Vogue Business, 22 May 2019, https://www.vogue.com/article/luxury-fashion-companies-charitable-donations

Boush, David M., Marian Friestad, and Gregory M. Rose. "Adolescent Skepticism toward TV Advertising and Knowledge of Advertiser Tactics." Journal of Consumer Research, vol. 21, no. 1, 1994, pp. 165–175. https://doi.org/10.2307/2489748

Darke, Peter R., and Robin J. B. Ritchie. "The Defensive Consumer: Advertising Deception, Defensive Processing, and Distrust." Journal of Marketing Research, vol. 44, no. 1, Feb. 2007, pp. 114–127. https://doi.org/10.1509/jmkr.44.1.114

Dean, Dwane Hal. "Consumer Perception of Corporate Donations: Effects of Company Reputation for Social Responsibility and Type of Donation." Journal of Advertising, vol. 32, no. 4, 2003, pp. 91–102. JSTOR, https://www.jstor.org/stable/4622182

Friestad, Marian, and Peter Wright. "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts." Journal of Consumer Research, vol. 21, no. 1, 1994, pp. 1–31. https://doi.org/10.1086/209380

Hamby, Anne, and David Brinberg. "Cause-Related Marketing Persuasion Knowledge: Measuring Consumers' Knowledge and Ability to Interpret CrM Promotions." Journal of Consumer Affairs, vol. 52, no. 2, 2018, pp. 373–392. https://doi.org/10.1111/joca.12167

Kuhn, Deanna. "Do Cognitive Changes Accompany Developments in the Adolescent Brain?" Perspectives on Psychological Science, vol. 1, no. 1, 2006, pp. 59–67. https://doi.org/10.1111/j.1745-6924.2006.00008.

Obermiller, Carl, and Eric R. Spangenberg. "Development of a Scale to Measure Consumer Skepticism toward Advertising." Journal of Consumer Psychology, vol. 7, no. 2, 1998, pp. 159–186. https://doi.org/10.1207/s15327663jcp0702_03

Parguel, Béatrice, Florence Benoît-Moreau, and Fabrice Larceneux. "How Sustainability Ratings Might Deter 'Greenwashing': A Closer Look at Ethical Corporate Communication." Journal of Business Ethics, vol. 102, no. 1, 2011, pp. 15–28. https://doi.org/10.1007/s10551-011-0901-2

Patagonia. "1% for the Planet." Patagonia, https://www.patagonia.com/one-percent-for-the-planet.html. Accessed 7 July 2026.

Pracejus, John W., G. Douglas Olsen, and Norman R. Brown. "On the Prevalence and Impact of Cause-Related Marketing Claims." Journal of Advertising, vol. 32, no. 3, 2003, pp. 19–30. https://doi.org/10.1080/00913367.2003.10639136

Steinberg, Laurence. "Cognitive and Affective Development in Adolescence." Trends in Cognitive Sciences, vol. 9, no. 2, 2005, pp. 69–74. https://doi.org/10.1016/j.tics.2004.12.00

Varadarajan, P. Rajan, and Anil Menon. “Cause-Related Marketing: A Coalignment of Marketing Strategy and Corporate Philanthropy.” Journal of Marketing, vol. 52, no. 3, 1988, pp. 58–74. JSTOR, https://doi.org/10.2307/1251450

AUTHORS SECTION

Misha Cherukumilli is a student at Gunn High School in Palo Alto, California. Her research focuses on consumer psychology and behavioral responses to CrM campaigns to further explore impacts of donation claims which she uses in her own personal business. She plans to pursue studies in business and economics.

-2.00 D: Rethinking ADHD as a Friction Between Mind and World

By Lila Kaya

Abstract

This essay argues that ADHD can be both over-diagnosed and better recognised, but that the familiar debate misses a deeper issue: ADHD may function less as a fixed disease category and more as a tool naming friction between certain minds and modern environments. Drawing on psychiatry, sociology, philosophy, and environmental health, the paper considers overdiagnosis, under-recognition in girls and adults, Ian Hacking's looping effect, Sara Shostak's socio-exposome, and John Locke's account of nominal essences. It concludes that ADHD names real struggles and can unlock necessary support, yet the illness frame may wrongly place responsibility only inside the individual rather than in the interaction between person and world.

Introduction

A myopic eye is not broken. It is an ordinary eye meeting a world built for sharp distance vision -- and the word “myopia” names the friction between an eye and its environment. What if some of what is now filed under mental illness works in a similar way: not as a defect inside the person, but as a difference that the modern world has turned into one?

The World Health Organization defines a mental disorder as “a clinically significant disturbance in cognition, emotional regulation, or behaviour,” and the U.S. National Institute of Mental Health lists ADHD among them (“Mental Disorders”; “Attention-Deficit/Hyperactivity Disorder”). Yet diagnoses of this genetically influenced “mental illness” climbed from 8 to 11 of every 100 American children in just eight years -- a 42 percent rise that Allen Frances described as “spreading like wildfire” (Visser et al.; Frances). That rise raises two natural questions. Has ADHD genuinely become more common, or are we simply over-diagnosing what was once treated as ordinary variation?

Both are partly right. Both are incomplete. Both share one unexamined assumption -- that ADHD is a real essence, a fixed object in nature that medicine either finds or misses. This essay challenges that assumption. As John Locke argued in 1689, some categories are not natural essences waiting to be discovered, but nominal essences: tools the mind uses to organise experience (Locke). ADHD, this essay argues, is one of those tools. It names real struggles and unlocks real support, but it should not be mistaken for the whole truth about what those struggles are or where they come from. Mental illness is too broad a category to defend or attack in a single essay, so the focus here is ADHD specifically, the most contested case of the wider debate.

Over-Diagnosis and Better Recognition

The over-diagnosis case is not a fringe view. Frances, who chaired the task force behind the DSM-IV, the American Psychiatric Association's diagnostic manual, now warns that diagnostic thresholds have drifted too far (Frances). Sami Timimi argues that ADHD is better understood as a cultural construct than a fixed biological entity, while Peter Conrad documents

how ordinary behaviour becomes medicalised (Timimi and Taylor; Conrad, The Medicalization of Society). A review of child and adolescent mental disorders likewise found evidence that overdiagnosis is a real concern (Merten et al.).

The numbers are striking. Between 1994 and 2009, stimulant prescriptions in the United States rose roughly sixfold, while DSM-5 widened eligibility for diagnosis (Paris et al.; American Psychiatric Association). A recent systematic review and meta-analysis found that children who were younger within their school year were 38 percent more likely to receive an ADHD diagnosis than their older classmates; notably, the relative-age effect appeared in teacher ratings of ADHD symptoms but not in parent ratings (Frisira et al.).

History deepens the suspicion. Stimulants were used on restless schoolchildren before the disorder had a name: their calming effect was observed in 1937, while the condition was named in 1957 and formalised in 1968 (Conrad, “The Discovery of Hyperkinesis”). This chronology does not prove that the diagnosis was built to fit the treatment; it does complicate the idea of ADHD as a natural object that was first discovered and only afterwards treated. Conrad and Bergey identify several forces that appear to have facilitated ADHD's international spread, including American psychiatry, pharmaceutical influence, looser DSM criteria, and advocacy groups. The category did not simply travel because it was discovered; it was also exported (Conrad and Bergey).

The opposite view is just as hard to ignore. ADHD was long missed in girls and women, whose symptoms often appear as quiet inattention rather than visible hyperactivity. Quinn and Madhoo describe how the prototype of the disruptive boy left many girls dismissed as “scattered” or “lazy.” Adult ADHD was also neglected because children were assumed to outgrow it (Conrad, The Medicalization of Society). Recognising these people was a correction, not an invention. The neurodiversity movement further reframed ADHD as part of naturally occurring cognitive variation rather than mere deficit (Singer; Silberman). For many, diagnosis after years of blame is relief.

So which is it -- over-diagnosis or better recognition? Both are happening, in different proportions for different groups. But stopping there leaves the most important part unexplained.

ADHD as a Category

Both answers quietly treat ADHD as a fixed object waiting to be found, like a tumour or a virus: either the diagnosis detects it correctly, or it mistakenly attaches it to someone who does not have it. But ADHD does not work that way. No clinical biomarker currently confirms it; diagnosis relies on symptom patterns, developmental history, functional impairment, and evidence that difficulties appear in more than one setting (“Attention-Deficit/Hyperactivity Disorder”; American Psychiatric Association). Reliability -- whether two clinicians applying the same checklist agree -- and validity -- whether the category names something real and stable in nature -- are not the same.

The philosopher Ian Hacking described why this matters with his idea of the “looping effect of human kinds” (Hacking). A tumour is an indifferent kind: it exists whether or not it is

named. ADHD is a human kind: the label changes the people it describes, and those people in turn reshape the label. The name even carries a hidden value judgment: Something is lacking. Something is disordered. On top of both, something is ill. Together, they sound like a person broken beyond fixing. A child whose attention keeps moving around the room is called "distractible"; a child who locks onto the worksheet and shuts the rest out is called "focused." Both are ways attention can work -- one that shifts, one that fixes. Only one is called a deficit. The deficit depends partly on what the surrounding world has decided to measure.
The looping effect also reframes "better recognition." Once ADHD became available as a framework, people could identify with it, seek help through it, and understand their past through it. That relief is genuine, but it is not the same as detecting a fixed object that was always waiting to be found. Suffering does not need a disease label to be legitimate. The category does not simply find its subjects; it also helps produce them. "Better recognition" cannot therefore be treated as a neutral measure, because the measuring instrument changes what it measures.

Environment and the Socio-Exposome

If the category is shaped by its surroundings, so is the behaviour beneath it. The sociologist Sara Shostak's concept of the socio-exposome holds that the environment is not only chemical but social -- schools, sleep, screens, stress -- and that all of these shape how genes express themselves (Senier et al.). Even where biology is real and inherited, genes alone rarely tell the whole story. Nerem and colleagues fed genetically near-identical rabbits the same high-cholesterol diet, yet one group developed sixty percent less arterial plaque -- the only difference being that a researcher had handled them affectionately. Same genes, same diet, different social world, different biology. This is not evidence about ADHD itself; it is evidence that social environment can affect biology (Nerem et al.).

The physician and author Thomas Armstrong makes a parallel case for ADHD, reframing it not as a definitive illness but as part of the natural diversity of the human brain (Armstrong). Studies have found associations among screen exposure, sleep problems, and ADHD symptoms, suggesting that these may be environmental factors intensifying the behaviours that attract diagnosis (Hong et al.; Winter and O'Neill). Professor Shostak notes that it is reasonable to ask both whether changes in the social environment alter gene expression and whether such changes might in turn increase the behaviors categorized as ADHD (Shostak). Professor Meredith Bergey adds in correspondence that the social environment may affect some behaviours independently of any link to neurobiology or genetics -- and that even if a physiological response were tied to the social environment, it need not be deemed a medical problem at all (Bergey). A child with the same neurological profile may function differently amid steady sleep and time outdoors than amid fragmented sleep and constant stimulation. The diagnostic question is incomplete without its companion: in what kind of environment?

Why does the medical label matter so much? Because labeling something as medical not only describes it -- it determines who must change. The sociologist Irving Zola called this the "depoliticization" of a problem: once a condition is deemed medical, the analytical lens shifts

entirely onto the individual, for both cause and cure, and the structural question quietly disappears (Zola). A pharmaceutical solution can be offered without redesigning classrooms, restructuring the school day, or revising screen policies -- only by treating the person who does not fit. The question of who built this world and for whom disappears within a medical frame—because the frame has already answered it: the problem is the individual. A diagnosis does not only describe a person. It also decides where responsibility falls.

Friction, Disability, and Treatment

If ADHD is friction with an environment rather than simply a defect inside a person, the same temperament may be disabling in one setting and useful in another. Attention that struggles through classroom stillness may better suit environments rewarding rapid scanning, quick reaction, novelty, and shifting focus (Stewman et al.). This is not the claim that ADHD is a "superpower": it can genuinely make school, time management, and emotional regulation harder. The point is that the deficit and asset depend partly on what the world asks of a person.
The strongest objection is severity. Some people with ADHD struggle at home, in friendships, and in ordinary daily routines, not only at school. That suffering should not be minimised. But severity proves that a difference can be disabling; it does not prove the difference is always an illness in itself -- any more than a predisposition to myopia makes short-sightedness a disease rather than a refractive error. The question is whether disability comes only from the brain, or from its collision with the world around it.

Medication seems to support the disease model because stimulants improve attention. But treatment working proves something about the treatment, not necessarily the nature of the condition. Caffeine sharpens a tired person without making tiredness a disease, and prescription stimulants improve focus in people without ADHD as well (Smith and Farah). Just as -2.00 D glasses correcting blurred vision do not prove the eye is diseased, a stimulant improving focus does not prove the brain is broken. None of this denies biology or disability. It argues that a difference can become disabling through the environment it must function within.

Conclusion

ADHD can therefore be both overdiagnosed and better recognised. Its boundary has widened under commercial, institutional, and educational pressure, while people once ignored -- especially girls and adults -- have gained language and support. But both trends grow from the same mistake: treating the diagnosis as a fixed essence rather than a tool naming friction between a mind and its world.

This matters in practice. Medication may help, but schools, screens, sleep, and daily life must also be reconsidered. In Locke's sense, ADHD is a nominal essence: a useful category that names real struggles and unlocks real support, but not the whole truth about why those struggles exist. If the label is mistaken for the truth itself, the harder question disappears: what kind of world has made these minds struggle?

Which returns us to the eye. A -2.00 D eye is not broken; it is an ordinary eye meeting a world built for sharp distance vision, and glasses reduce the friction without turning the eye into a disease. Perhaps ADHD should be approached with the same humility: not as a fixed defect inside the person, but as a real human difference whose harm depends partly on the world around it. The question is not whether the struggles are real -- they are -- but what kind of world makes them disabling, and whether we are brave enough to change that world rather than only the people who struggle within it.

Note

Personal communications are cited with permission; the correspondence is on file and available on request.

Acknowledgements

I am sincerely grateful to Professor Sara Shostak and Professor Meredith Bergey for generously giving their time and thought to my questions, and for their willingness to engage so openly with a student's work.

Works Cited

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed., American Psychiatric Publishing, 2013.

Armstrong, Thomas. Neurodiversity: Discovering the Extraordinary Gifts of Autism, ADHD, Dyslexia, and Other Brain Differences. Da Capo Lifelong, 2010.

Bergey, Meredith R. Email to the author. 12 May 2026 and 16 May 2026.

Conrad, Peter. "The Discovery of Hyperkinesis: Notes on the Medicalization of Deviant Behavior." Social Problems, vol. 23, no. 1, 1975, pp. 12-21.

---. The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders. Johns Hopkins UP, 2007.

Conrad, Peter, and Meredith R. Bergey. "The Impending Globalization of ADHD: Notes on the Expansion and Growth of a Medicalized Disorder." Social Science & Medicine, vol. 122, 2014, pp. 31-43.

Frances, Allen. Saving Normal: An Insider's Revolt against Out-of-Control Psychiatric Diagnosis, DSM-5, Big Pharma, and the Medicalization of Ordinary Life. William Morrow, 2013.

Frisira, E., J. Holland, and K. Sayal. "Systematic Review and Meta-Analysis: Relative Age in Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder." European Child & Adolescent Psychiatry, vol. 34, no. 2, 2025, pp. 381-401. https://doi.org/10.1007/s00787-024-02459-x.

Hacking, Ian. The Social Construction of What? Harvard UP, 1999.

Hong, George C. C., et al. "Diet, Physical Activity, and Screen Time to Sleep Better: Multiple Mediation Analysis of Lifestyle Factors in School-Aged Children with and without Attention Deficit Hyperactivity Disorder." Journal of Attention Disorders, vol. 25, no. 13, 2021, pp. 1847-1858. https://doi.org/10.1177/1087054720940417.

Locke, John. An Essay Concerning Human Understanding. 1689.

Merten, Eva C., et al. "Overdiagnosis of Mental Disorders in Children and Adolescents in Developed Countries." Child and Adolescent Psychiatry and Mental Health, vol. 11, 2017, article 5.

National Institute of Mental Health. "Attention-Deficit/Hyperactivity Disorder." National Institute of Mental Health, www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd.

Nerem, Robert M., Maurice J. Levesque, and J. Fred Cornhill. "Social Environment as a Factor in Diet-Induced Atherosclerosis." Science, vol. 208, no. 4451, 1980, pp. 1475-1476.

Paris, Joel, Vikram Bhat, and Brett Thombs. "Is Adult Attention-Deficit Hyperactivity Disorder Being Overdiagnosed?" Canadian Journal of Psychiatry, vol. 60, no. 7, 2015, pp. 324-328.

Quinn, Patricia O., and Manisha Madhoo. "A Review of Attention-Deficit/Hyperactivity Disorder in Women and Girls: Uncovering This Hidden Diagnosis." The Primary Care

Companion for CNS Disorders, vol. 16, no. 3, 2014. https://doi.org/10.4088/PCC.13r01596.

Senier, Laura, et al. “The Socio-Exposome: Advancing Exposure Science and Environmental Justice in a Post-Genomic Era.” Environmental Sociology, vol. 3, no. 2, 2017, pp. 107-121. https://doi.org/10.1080/23251042.2016.1220848.

Shostak, Sara. Email to the author. 11 May 2026.

Silberman, Steve. NeuroTribes: The Legacy of Autism and the Future of Neurodiversity. Avery, 2015.

Singer, Judy. Odd People In: The Birth of Community amongst People on the Autism Spectrum. Honours thesis, University of Technology Sydney, 1998.

Smith, M. Elizabeth, and Martha J. Farah. “Are Prescription Stimulants ‘Smart Pills’? The Epidemiology and Cognitive Neuroscience of Prescription Stimulant Use by Normal Healthy Individuals.” Psychological Bulletin, vol. 137, no. 5, 2011, pp. 717-741. https://doi.org/10.1037/a0023825.

Stewman, Cameron G., et al. “Attention Deficit Hyperactivity Disorder: Unique Considerations in Athletes.” Sports Health, vol. 10, no. 1, 2018, pp. 40-46. https://doi.org/10.1177/1941738117742906.

Timimi, Sami, and Eric Taylor. “ADHD Is Best Understood as a Cultural Construct.” The British Journal of Psychiatry, vol. 184, no. 1, 2004, pp. 8-9.

Visser, Susanna N., et al. “Trends in the Parent-Report of Health Care Provider-Diagnosed and Medicated Attention-Deficit/Hyperactivity Disorder: United States, 2003-2011.” Journal of the American Academy of Child & Adolescent Psychiatry, vol. 53, no. 1, 2014, pp. 34-46.

Winter, H., and J. O'Neill. “The Impact of Screen Time on ADHD Symptoms in Children and Adolescents: A Narrative Review of Treatment Approaches.” Clinical Child Psychology and Psychiatry, 2026. https://doi.org/10.1177/13591045251399697.

World Health Organization. “Mental Disorders.” World Health Organization, 8 June 2022, www.who.int/news-room/fact-sheets/detail/mental-disorders.

Zola, Irving Kenneth. “Medicine as an Institution of Social Control.” The Sociological Review, vol. 20, no. 4, 1972, pp. 487-504.

A Comparative Study of Customer Satisfaction and Consumption Behavior Across Three Shopping Centers in Shanghai: Evidence from a Questionnaire Survey By Kaiqi Li

1. Introduction

1.1 Research Background and Problem Statement

In the context of the experience economy and the ongoing transformation of the retail industry, shopping centers are increasingly competing through differentiated positioning rather than homogeneous expansion. Customer satisfaction has long been recognized as a key determinant of customer loyalty and business performance (Oliver, 1980; Fornell et al., 1996). As consumer expectations continue to evolve, understanding how different positioning strategies influence customer profiles, consumption behavior, and satisfaction has become increasingly important for both researchers and practitioners.

To explore these issues, this study examines three shopping centers in Shanghai with distinct market positioning: Huijin Department Store ("Huijin"; family life center), Grand Gateway 66 ("Ganghui"; luxury fashion destination), and Metro City ("Meiluo"; youth culture hub). Based on a questionnaire survey of 116 valid respondents, the study investigates how differentiated positioning shapes customer characteristics, consumption behavior, and multidimensional customer satisfaction.

Despite the widespread adoption of differentiated positioning strategies, an important question remains: do these strategies translate into measurable differences in customer characteristics, consumption behavior, and customer satisfaction? While shopping centers often promote distinct market positioning, empirical comparisons of their effectiveness remain relatively limited, particularly within the same urban commercial environment. To address this issue, this study compares three shopping centers in Xujiahui, Shanghai, each representing a different market positioning strategy.

Although differentiated positioning has become a widely adopted strategy in the retail industry, its effectiveness has not been fully examined through empirical comparisons. Existing studies have primarily focused on customer satisfaction or consumer behavior in individual retail settings, while relatively few have investigated how different positioning strategies shape customer profiles, consumption behaviors, and satisfaction within the same commercial district. As a result, the relationships among shopping center positioning, customer characteristics, consumption behavior, and customer satisfaction remain insufficiently understood. This study addresses this research gap through a comparative analysis of three shopping centers in Shanghai.

To address this gap, this study examines three shopping centers located in Xujiahui, Shanghai, one of the city's major commercial districts. Despite their close geographic proximity, the three shopping centers have adopted distinct market positioning strategies. Huijin Department Store serves primarily as a family-oriented shopping destination, Grand Gateway 66 focuses on luxury fashion and lifestyle consumption, and Metro City targets younger consumers

through cultural, entertainment, and technology-related retail offerings. Their contrasting positioning strategies provide an appropriate setting for examining how market positioning influences customer demographics, consumption behavior, and customer satisfaction.

1.2 Research Significance, Objectives, and Questions

This research holds significant theoretical importance and practical value.

At the theoretical level, the study combines the classic customer satisfaction theory (the ACSI model) with market segmentation and consumer behavior theories and applies them to the comparative analysis of specific commercial entities in China's first-tier cities. This helps to enrich and develop localized retail and consumer behavior theories.

At the practical level, the findings of this research can provide direct decision-making references for the operation managers, investors, and commercial real estate planners of shopping centers. By quantitatively presenting the effectiveness of different positioning strategies, consumption contributions, and satisfaction feedback, it can help managers more objectively evaluate the effectiveness of their positioning and identify strengths and weaknesses in the service process, thereby optimizing operational resource allocation, formulating more targeted marketing strategies, and upgrading experience plans, ultimately achieving a dual improvement in customer satisfaction and business value.

Accordingly, this study addresses the following research questions:

RQ1: How do the customer demographic characteristics (e.g., age, gender, and occupation) and consumption behaviors differ across the three shopping centers with distinct market positioning?

RQ2: How do customer satisfaction levels differ among the three shopping centers in terms of overall satisfaction and specific service dimensions, including transportation convenience, environment, cleanliness, parking, retail mix, service quality, price, and promotional activities?

RQ3: Which demographic characteristics and consumption behavior factors are significantly associated with higher overall customer satisfaction, and what managerial implications can be drawn from these relationships?

To answer these questions, this study compares customer profiles, consumption behaviors, and satisfaction across three representative shopping centers in Shanghai using questionnaire survey data and statistical analyses.

2. Literature Review and Theoretical Framework

2.1 Customer Satisfaction in Retail Settings

Customer satisfaction has long been regarded as a fundamental indicator of business performance and customer loyalty. The Expectation–Confirmation Theory suggests that satisfaction is formed by comparing customers' expectations with their actual consumption

experiences (Oliver, 1980). Building upon this framework, the American Customer Satisfaction Index (ACSI) proposes that customer satisfaction is jointly influenced by customer expectations, perceived quality, and perceived value, and subsequently affects customer loyalty and future purchasing behavior (Fornell et al., 1996; Fornell, 2021).

In the retail industry, customer satisfaction is widely recognized as a multidimensional concept rather than a single evaluation. Previous studies have shown that factors such as accessibility, physical environment, cleanliness, service quality, product variety, price fairness, and promotional activities all contribute to customers' overall shopping experience (Parasuraman et al., 1988; Baker et al., 2002; Kotler & Keller, 2016). Accordingly, this study evaluates customer satisfaction across nine dimensions: transportation convenience, environmental comfort, hygiene, parking convenience, retail mix diversity, service quality, price rationality, promotional intensity, and overall satisfaction.

2.2 Demographic Characteristics and Customer Satisfaction

Previous research suggests that customer satisfaction varies across demographic groups because customers differ in their expectations, purchasing power, and shopping preferences. Age is one of the most influential factors. Younger consumers generally value entertainment, fashion, social interaction, and digital experiences, whereas middle-aged and older consumers tend to emphasize convenience, service quality, comfort, and family-oriented facilities (Arnold & Reynolds, 2003; Kotler & Keller, 2016).

Gender and occupation also influence shopping preferences. Female consumers are often more sensitive to service quality, shopping atmosphere, and product variety, while occupational status is closely related to disposable income, shopping frequency, and consumption purposes. Students, office workers, and retirees may therefore evaluate the same shopping environment differently because of their distinct consumption needs and lifestyles (Solomon, 2020).

These findings suggest that demographic characteristics should be considered when evaluating customer satisfaction in shopping centers with different market positioning.

2.3 Consumption Behavior and Customer Satisfaction

Consumption behavior provides an important link between customer characteristics and satisfaction. Shopping frequency reflects customer engagement and familiarity with a shopping center. Frequent visitors often develop stronger loyalty, although repeated visits may also increase expectations regarding service quality (Fornell et al., 1996).

Spending level is another important behavioral indicator. Customers with different expenditure levels may perceive value differently. Moderate spending often represents a balance between monetary cost and perceived benefits, while very high expenditures may be accompanied by higher expectations, making customer satisfaction more difficult to achieve (Kotler & Keller, 2016).

Consumption purpose also shapes customer evaluations. Customers visiting primarily for dining, family activities, shopping, entertainment, or cultural experiences tend to focus on

different aspects of the shopping environment. Therefore, understanding consumption behavior is essential for explaining differences in customer satisfaction among shopping centers with distinct positioning.

2.4 Research Gap and Theoretical Framework

Although previous studies have extensively examined customer satisfaction and consumer behavior, most have focused on general retail environments or single shopping centers. Comparative studies investigating how different positioning strategies influence customer profiles, consumption behaviors, and satisfaction within the same urban commercial district remain relatively limited.

To address this gap, this study compares three shopping centers with distinct market positioning in Shanghai. Based on existing customer satisfaction and consumer behavior theories, an integrated analytical framework is developed in which shopping center positioning influences customer demographic characteristics, which in turn shape consumption behaviors and ultimately affect multidimensional customer satisfaction. This framework provides the theoretical foundation for the empirical analyses presented in the following sections.

3. Methods

3.1 Overall Research Design

This study employed a quantitative, cross-sectional research design to examine customer satisfaction and consumption behavior across three shopping centers in Shanghai. Data were collected through a structured questionnaire administered to customers at the selected shopping centers during a single survey period. The questionnaire gathered information on respondents' demographic characteristics, consumption behaviors, and evaluations of multiple dimensions of customer satisfaction. The collected data were subsequently analyzed using descriptive statistics, analysis of variance (ANOVA), and logistic regression to identify differences among shopping centers and explore factors associated with customer satisfaction.

3.2 Data Collection and Samples

The data were collected in the summer of 2025 in Shanghai. I collaborated with assigned surveyors at the three shopping centers and randomly invited customers to participate voluntarily in anonymous questionnaire surveys at the main entrances, rest areas, and popular brand stores of these shopping centers on both weekdays and weekends and during different periods of the day (morning, afternoon, and evening).. A total of 121 questionnaires were distributed, and after excluding those with incomplete answers or obvious logical contradictions, 116 valid questionnaires were retained for analysis, yielding a valid response rate of 95.9%.

The final sample included 17 respondents from Huijin Department Store (14.7%), 50 from Grand Gateway 66 (43.1%), and 49 from Metro City (42.2%). Although the sample sizes were not evenly distributed across the three shopping centers, each shopping center represented a

distinct market positioning and customer group, making the sample appropriate for comparative analysis. This approach is consistent with cross-sectional studies that have examined demographic differences in mall shopping satisfaction across diverse retail settings (Jones & Lee, 2022).

3.3 Questionnaire Measures

The questionnaire consisted of four sections designed to collect information on respondents' demographic characteristics, consumption behavior, customer satisfaction, and qualitative feedback.

Demographic information: Respondents were asked to report their age, gender, and occupation. Age was classified into three groups: 35 years or younger, 36–55 years, and 56 years or older.

Consumption behavior: Questions measured respondents' shopping frequency, typical spending per visit, and primary consumption purposes. Shopping frequency was classified into five categories (weekly, monthly, quarterly, semi-annually, and annually). Spending per visit was categorized into five ranges (<100 RMB, 100–500 RMB, 500–1000 RMB, 1000–5000 RMB, and >5000 RMB). Respondents were also asked to identify their primary purposes for visiting the shopping center, including dining, family activities, brand events, new product experiences, cultural activities, and personal shopping.

Customer satisfaction: Customer satisfaction was measured using a 10-point Likert scale, where 1 represented "very dissatisfied" and 10 represented "very satisfied." Respondents evaluated their most recent shopping experience across nine dimensions, including transportation convenience, environmental comfort, cleanliness, parking convenience, retail mix diversity, service quality, price rationality, promotional activities, and overall satisfaction. The use of a 10-point scale allows for greater differentiation in respondents' evaluations than shorter rating scales.

Open-ended feedback: Respondents were invited to provide comments and suggestions regarding their shopping experiences. These qualitative responses were used to supplement the quantitative findings and provide additional insights into customer perceptions.

3.4 Data Analysis Methods

All statistical analyses were performed using R, primarily with the **tidyverse** (Wickham, 2022), **ggplot2** (ggplot2 Team, 2023), and **car** packages.

Descriptive statistical analysis: Frequencies, percentages, means, and standard deviations were calculated to summarize respondents' demographic characteristics, consumption behaviors, and customer satisfaction scores. **Comparative analysis**: One-way analysis of variance (ANOVA) was conducted to examine differences in customer satisfaction among the three shopping centers. If the results are significant, further comparisons will be conducted to clarify the specific sources of the differences. **Logistic regression analysis**: Binary logistic regression was conducted to identify demographic and consumption behavior factors associated

with high customer satisfaction. Regression coefficients, odds ratios (ORs), 95% confidence intervals, and p-values were reported to evaluate the strength and statistical significance of each predictor. Statistical significance was evaluated at the 0.05 level.

4. Results

4.1 User Profile: Highly Differentiated Customer Segments

Table 1 presents the demographic characteristics of respondents across the three shopping centers. Significant differences were observed in age, gender, and occupation among the three shopping centers (Chi-square tests, $p < 0.05$). Overall, Huijin Department Store had a more balanced age and gender distribution, Grand Gateway 66 attracted a relatively younger customer base, and Metro City was dominated by young student customers.

Table 1. Demographic Characteristics of Respondents Across the Three Shopping Centers

Characteristic	Classification	Total (n=116)	Huijin Department Store (n=17)	Grand Gateway 66 (n=50)	Metro City (n=49)
Age	≤ Age of 35	58.6%	23.5%	70.0%	93.9%
	Age of 36-55	32.8%	41.2%	26.0%	4.1%
	≥ Age of56	8.6%	35.3%	4.0%	2.0%
Gender	Female	62.9%	47.1%	56.0%	81.6%
	Male	37.1%	52.9%	44.0%	18.4%
Occupation	Student	48.3%	11.8%	34.0%	77.6%
	Office Worker	38.8%	52.9%	46.0%	18.4%
	Retiree	7.8%	29.4%	4.0%	0.0%
	Others	5.1%	5.9%	16.0%	4.0%

Age distribution: Age distribution differed considerably across the three shopping centers. Respondents aged 36–55 years accounted for the largest proportion of customers at Huijin Department Store (41.2%), whereas respondents aged 35 years or younger represented the majority of customers at Grand Gateway 66 (70.0%) and Metro City (93.9%).

Gender composition: Gender composition also varied among the three shopping centers. Huijin Department Store had a relatively balanced gender distribution (52.9% male and 47.1%

female), whereas female respondents accounted for 56.0% of customers at Grand Gateway 66 and 81.6% at Metro City.

Occupational distribution: Occupation distributions also differed across shopping centers. Office workers constituted the largest customer group at Huijin Department Store (52.9%), while students represented the largest proportion of respondents at Metro City (77.6%). Grand Gateway 66 had a more mixed occupational composition, with office workers (46.0%) and students (34.0%) accounting for the largest shares of respondents.

4.2 Consumption Behavior Across the Three Shopping Centers

Table 2 presents respondents' consumption behavior across the three shopping centers. Differences were observed in spending patterns and primary consumption purposes, whereas shopping frequency showed relatively similar distributions across the three shopping centers.

Table 2. Consumption Behavior of Respondents Across the Three Shopping **Centers**

Behavior	Category	Huijin Department Store	Grand Gateway 66	Metro City
Frequency of consumption	Per week	29.4%	34.0%	24.5%
	Per month	41.2%	42.0%	42.9%
	Quarterly / less frequently	29.4%	24.0%	32.6%
Consumption amount	<100 RMB	0.0%	8.0%	20.4%
	100-500 RMB	41.2%	54.0%	61.2%
	500-1000 RMB	41.2%	28.0%	14.3%
	1000-5000RMB	11.8%	8.0%	4.1%
	>5000 RMB	5.8%	2.0%	0.0%
Main consumption purpose	Catering	29.4%	46.0%	32.7%
	Family/Child Needs	23.5%	12.0%	8.2%
	Brand Events	11.8%	16.0%	12.2%
	New Product Experience	5.9%	10.0%	16.3%

Behavior	Category	Huijin Department Store	Grand Gateway 66	Metro City
	Cultural Activities	5.9%	4.0%	14.3%

Frequency of consumption: Shopping frequency showed similar distributions across the three shopping centers. Monthly visits were the most common frequency, accounting for approximately 41%–43% of respondents in each shopping center.

Consumption amount: Respondents at Huijin Department Store showed relatively higher spending levels, with 41.2% spending 100–500 RMB and another 41.2% spending 500–1000 RMB per visit. At Grand Gateway 66, most respondents (54.0%) reported spending between 100 and 500 RMB. Metro City had the largest proportion of respondents spending less than 100 RMB (20.4%) and the highest concentration in the 100–500 RMB category (61.2%).

Consumption purpose: Dining was the most frequently reported consumption purpose across all three shopping centers. However, differences were observed in secondary consumption purposes. Family-related activities accounted for a larger share of visits to Huijin Department Store (23.5%), brand events were more common at Grand Gateway 66 (16.0%), and new product experiences (16.3%) and cultural activities (14.3%) were reported more frequently at Metro City.

4.3 Comparison of Customer Satisfaction

Table 3 presents the mean customer satisfaction scores across the three shopping centers. One-way ANOVA indicated significant differences in transportation convenience, environmental comfort, cleanliness, parking convenience, and promotional intensity ($p < 0.05$), whereas differences in overall satisfaction, retail mix diversity, service quality, and price rationality were not statistically significant.

Table 3. Mean Customer Satisfaction Ratings Across the Three Shopping Centers

Satisfaction dimensions	Overall mean	Huijin Department Store	Grand Gateway 66	Metro City	F value	p value
Overall satisfaction	8.05	8.18	8.13	7.87	2.98	0.054
Transportation convenience	8.95	9.24	8.82	8.86	4.02	0.020*
Environmental comfort	8.58	8.65	8.72	8.39	3.15	0.046*
Sanitary cleanliness	8.62	9.12	8.88	8.00	16.34	<0.001***

Satisfaction dimensions	Overall mean	Huijin Department Store	Grand Gateway 66	Metro City	F value	p value
Parking convenience	7.35	6.76	7.86	7.29	5.67	0.004**
Retail Mix Diversity	8.19	8.24	8.20	8.14	0.17	0.847
Service quality	8.31	8.12	8.46	8.29	1.45	0.238
Price rationality	7.39	7.71	7.32	7.22	1.82	0.166
Discount promotion intensity	6.99	7.59	6.78	6.76	4.68	0.011*

*Note: $p < 0.05$, **$p < 0.01$, **$p < 0.001$*

Huijin Department Store received the highest mean scores for transportation convenience (9.24), cleanliness (9.12), price rationality (7.71), promotional intensity (7.59), and overall satisfaction (8.18). **Grand Gateway 66** achieved the highest ratings for environmental comfort, service quality, and parking convenience.

Metro City recorded the lowest overall satisfaction score and the lowest ratings for cleanliness and environmental comfort.

Parking convenience received the lowest mean rating across all three shopping centers.

4.4 Factors Associated with High Customer Satisfaction

To identify demographic and consumption behavior factors associated with high customer satisfaction, descriptive group comparisons and binary logistic regression analyses were conducted.

Descriptive group comparisons: Respondents aged 36–55 years reported the highest mean overall satisfaction score (8.46), compared with respondents aged 35 years or younger (7.94) and those aged 56 years or older (7.81). Across spending categories, respondents spending 500–1000 RMB per visit reported the highest mean satisfaction score (8.58), whereas those spending 1000–5000 RMB reported the lowest mean score (7.06). Respondents spending more than 5000 RMB also reported relatively high satisfaction (9.00), although this category included only a small number of respondents.

Binary logistic regression: To account for the simultaneous effects of multiple variables, a binary logistic regression model was estimated (Table 4). The results are presented below.

Table 4. Binary Logistic Regression Results for High Customer Satisfaction (≥ 8 Points)

Predictor variable	β coefficient	Standard error	Z value	p value	OR (Odds Ratio)	95% CI
(Intercept)	-0.8912	0.301	-2.961	0.003**	0.410	[0.228, 0.739]
Age group: age of 36 - 55	1.3727	0.412	3.331	0.001*	3.947	[1.761, 8.843]
Age group: ≥ age of 56	0.4561	0.538	0.848	0.396	1.578	[0.550, 4.527]
Consumption amount: < 100 RMB	0.2105	0.489	0.430	0.667	1.234	[0.474, 3.215]
Consumption amount: 500 - 1000 RMB	0.8943	0.385	2.323	0.020*	2.445	[1.150, 5.199]
Consumption amount: 1000 - 5000 RMB	-1.1021	0.601	-1.833	0.067	0.332	[0.102, 1.078]
Consumption amount: > 5000 RMB	1.5432	1.203	1.283	0.199	4.679	[0.446, 49.084]

** The reference categories were respondents aged 35 years or younger and those spending 100–500 RMB per visit. Model $\chi^2 = 18.34$, $p < 0.01$; Nagelkerke $R^2 = 0.21$.*

The results of the binary logistic regression analysis are presented in Table 4. After controlling for demographic and consumption behavior variables, age and spending level were significantly associated with the likelihood of reporting high customer satisfaction.

Age: Respondents aged 36–55 years were significantly more likely to report high customer satisfaction than respondents aged 35 years or younger (OR = 3.95, 95% CI = 1.76–8.84, $p = 0.001$).

Spending level: Respondents spending 500–1000 RMB per visit were significantly more likely to report high customer satisfaction than those spending 100–500 RMB (OR = 2.45, 95% CI = 1.15–5.20, $p = 0.020$).

Other variables: No statistically significant associations were observed for respondents aged 56 years or older or for the remaining spending categories.

5. Discussions

5.1 Implications of Differentiated Market Positioning

The findings of this study suggest that the three shopping centers exhibited distinct customer profiles, consumption behaviors, and customer satisfaction patterns that were generally

consistent with their respective market positioning. Huijin Department Store attracted a relatively balanced customer population with a higher proportion of middle-aged and retired customers, Grand Gateway 66 primarily attracted younger office workers and fashion-oriented consumers, and Metro City was dominated by younger customers, particularly students. These differences indicate that shopping centers with different positioning strategies may attract different customer segments and provide distinct shopping experiences.

The customer satisfaction results further suggest that different shopping centers demonstrated strengths in different dimensions. Huijin Department Store received relatively high ratings for transportation convenience and cleanliness, Grand Gateway 66 performed well in environmental comfort and service quality, while Metro City showed comparatively lower ratings in several satisfaction dimensions. Rather than indicating that one shopping center was universally superior, these findings suggest that customer satisfaction may depend on the extent to which a shopping center's services and environment meet the expectations and preferences of its primary customer groups. This interpretation is generally consistent with previous studies emphasizing the importance of perceived quality and customer value in shaping customer satisfaction (Fornell et al., 1996; Kotler & Keller, 2016).

At the same time, the findings also highlight potential challenges associated with highly differentiated positioning. Metro City, while attracting a predominantly young customer base, received relatively lower ratings for environmental comfort and cleanliness. Although this study cannot establish causal relationships, the results suggest that maintaining high service quality remains important regardless of market positioning. Differentiated positioning may help attract target customer groups, but it should complement rather than replace the provision of a consistently satisfactory shopping environment.

5.2 Implications of Customer Age and Spending Level

The logistic regression analysis identified customer age and spending level as two factors associated with higher customer satisfaction. Respondents aged 36–55 years and those spending between 500 and 1000 RMB per visit were more likely to report high satisfaction than the corresponding reference groups. Although these findings do not imply causal relationships, they suggest that customer characteristics and spending behavior may influence shopping experiences and satisfaction evaluations.

The relatively high satisfaction reported by respondents aged 36–55 years may reflect the extent to which the shopping centers meet the needs of economically active consumers with stable purchasing power and diverse shopping purposes. This age group often combines shopping with dining, family activities, and leisure consumption, making them an important customer segment for shopping center operators. Therefore, maintaining a balanced mix of retail, dining, and family-oriented services may help improve the shopping experience for this group.

Similarly, respondents spending between 500 and 1000 RMB per visit were more likely to report high satisfaction than those in the reference spending category. One possible explanation is that this spending range may represent a balance between perceived value and

shopping experience. However, this interpretation should be treated with caution because the present study cannot determine the underlying reasons for customers' spending decisions. Future research using larger samples or qualitative interviews could further examine the relationship between spending level and customer satisfaction.

Overall, these findings suggest that shopping center managers may benefit from tailoring retail offerings and service strategies to the characteristics of their primary customer groups while maintaining a high-quality shopping environment for customers with different spending levels.

5.3 Parking Convenience as a Common Area for Improvement

One notable finding of this study is that parking convenience received the lowest satisfaction ratings across all three shopping centers. Although the absolute scores differed slightly, parking consistently received lower evaluations than the other satisfaction dimensions, suggesting that it represents a common concern among customers regardless of shopping center positioning.

Several factors may contribute to relatively low parking satisfaction, such as limited parking availability, parking fees, or difficulties in accessing parking facilities (Wang et al., 2024). However, the present study did not investigate the specific causes of these evaluations, and therefore no causal conclusions can be drawn. Future research could combine questionnaire surveys with observational data or customer interviews to better understand the factors affecting parking satisfaction.

From a practical perspective, shopping center managers may consider evaluating existing parking services and identifying opportunities to improve customer convenience. Enhancing parking information, improving wayfinding, or optimizing parking management may contribute to improving the overall shopping experience. Nevertheless, the effectiveness of these measures requires further investigation.

5.4 Practical Implications

The findings of this study provide several practical implications for shopping center management. First, customer profiles differed across the three shopping centers, suggesting that management strategies should be aligned with the characteristics and preferences of their primary customer groups. Rather than adopting uniform marketing approaches, shopping centers may benefit from developing services and retail offerings that better match the needs of their target customers. In this regard, the integration of online and offline channels has emerged as an increasingly important strategy for enhancing customer satisfaction in retail malls (Zhang, 2023).

Second, the results indicate that customer satisfaction was associated with multiple dimensions of the shopping experience, including transportation convenience, environmental comfort, cleanliness, and service quality. While different shopping centers demonstrated strengths in different areas, maintaining consistent service quality across these dimensions may contribute to higher overall customer satisfaction.

Finally, parking convenience received relatively low satisfaction ratings across all three shopping centers. Although this study did not investigate the underlying causes of this finding, it suggests that improving customer convenience beyond the retail environment may represent an opportunity for enhancing the overall shopping experience. Regular customer feedback and continuous monitoring of service quality may help shopping center managers identify areas requiring further improvement.

6. Conclusion

This study compared customer satisfaction and consumption behavior across three shopping centers with different market positioning in Shanghai. The findings indicate that the three shopping centers attracted different customer groups and exhibited distinct patterns of consumption behavior and customer satisfaction. In particular, respondents aged 36–55 years and those spending between 500 and 1000 RMB per visit were more likely to report high customer satisfaction, while parking convenience consistently received the lowest satisfaction ratings across all three shopping centers.

Overall, the results suggest that customer satisfaction is associated with both customer characteristics and shopping experiences. These findings align with the broader theoretical understanding that customer satisfaction frameworks continue to evolve in response to changing market dynamics (Fornell, 2021). Although the study cannot establish causal relationships, it provides a comparative perspective on how different market positioning strategies may be reflected in customer profiles and satisfaction outcomes. These findings may offer useful insights for shopping center managers seeking to better understand customer needs and improve customer experience.

This study has several limitations. First, the sample size was relatively small, particularly for Huijin Department Store, which may limit the generalizability of the findings. Second, the cross-sectional research design does not allow causal relationships to be established. Third, this study focused primarily on demographic characteristics and consumption behaviors, while other factors, such as customer expectations, brand perception, and emotional responses, were not examined.

Future research could employ larger and more representative samples, incorporate additional psychological and behavioral variables, and apply longitudinal or mixed-method research designs to further examine the relationships among shopping center positioning, customer behavior, and customer satisfaction.

Works Cited

Arnold, M. J., & Reynolds, K. E. (2003). Hedonic shopping motivations. Journal of Retailing, 79(2), 77–95.

Baker, J., Parasuraman, A., Grewal, D., & Voss, G. B. (2002). The influence of multiple store environment cues on perceived merchandise value and patronage intentions. Journal of Marketing, 66(2), 120–141.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American Customer Satisfaction Index: Nature, purpose, and findings. Journal of Marketing, 60(4), 7–18.

Fornell, C. (2021). The American Customer Satisfaction Index: Nature, Purpose, and Findings. Journal of Economic Perspectives, 17(1), 93–110.

Jones, A., & Lee, S. (2022). Demographic Differences in Mall Shopping Satisfaction: A Cross-Sectional Study. Journal of Retailing and Consumer Services, 66, 102889.

Kotler, P., & Keller, K. L. (2016). Marketing Management (15th ed.). Pearson.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460–469.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12–40.

Solomon, M. R. (2020). Consumer Behavior: Buying, Having, and Being (13th ed.). Pearson.

Wang, L., et al. (2024). Parking Satisfaction and Customer Loyalty in Shopping Malls: The Mediating Role of Overall Experience. Sustainability, 16(3), 1245.

Wickham, H. (2022). tidyverse: Easily Install and Load the Tidyverse. R package version 1.3.2. https://CRAN.R-project.org/package=tidyverse

ggplot2 Team. (2023). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R package version 3.4.1. https://ggplot2.tidyverse.org/

Zhang, H. (2023). Online-Offline Integration Mode and Customer Satisfaction in Retail Malls. Electronic Commerce Research and Applications, 56, 101234.

Evaluating AI Trading Bots: A Comparative Study of Reinforcement-Learning Strategies

By Navya Bhandari

Abstract

The use of reinforcement learning-based AI in financial markets has been rising over the years. The market has been saturated with many different platforms and strategies, making it important to understand their effectiveness. The objective of this study is to evaluate the performance of multiple reinforcement-learning (RL) trading bots (A2C, DDPG, PPO, TD3, SAC) against benchmarks such as the Mean-Variance strategy and DJIA Buy-and-Hold. Additionally, another goal is to assess the impact of modifying key parameters (transaction cost, volatility, gate, HMax) on performance. In order to achieve this, bots in the open-source FinRL repository were used. Backtests were run on unmodified and modified versions of code to analyze the efficiency of each bot as well as the impact of changed parameters. Through the data, it was deduced that A2C outperformed all other bots and displayed the most consistency across all metrics. Although Mean-Variance slightly outperformed A2C in raw return, it lacked consistency. Additionally, small parameter tweaks (transaction costs and HMax) had minimal effects on bot performance. However, volatility gate adjustments had significant impacts, with reduced gates causing substantial performance declines across all bots. This identified market stability as a critical factor in bot effectiveness. Overall, reinforcement learning bots such as A2C show strong potential for trading. Future research can expand to different market conditions, longer simulations, adaptation to market state, etc. These bots and related studies can be used for potential applications such as financial inclusion tools, cooperative fund management, and educational platforms.

Introduction

What are Trading Bots? Trading bots are software tools commonly used in cryptocurrencies, stocks, and commodities. They automatically execute trades for users based on predefined algorithms and strategies. They analyze various market data, identify trading opportunities, and react to market movements instantly. This makes these bots particularly popular in the volatile cryptocurrency market, but their use case is universal. This paper focuses on a set of bots based on the stock market. Such bots can be purchased through various online platforms, sometimes even offered by brokers for free (Global Finance Trading, 2023). Their profitability often depends on market conditions, trading strategies, and their algorithm.

Trading bots are completely legal, but like most automation tools, have a set of rules and regulations to follow (LegalClarity, 2025). These are some illegal practices:

1) Spoofing: placing large bids just to withdraw from them right before execution. This creates false impressions of market demand.
2) Wash trading: simultaneously buying and selling the same financial instrument. This distorts true market prices.

3) Front-running: Using advanced, non-public knowledge of future transactions for personal benefit.

Auto trading occurs when trading bots combine data analysis and programming to make real-time decisions. This entails four main steps:

1) Data Analysis: The software continuously monitors the market, tracking changes in prices, trading activity, and indicator signals. This information guides its trading decisions.
2) Algorithmic Decisions: Based on its programmed algorithms, the bot decides when to enter or exit trades, following strategies that match the trader's desired risk and return balance.
3) Trade Execution: After deciding the best course of action, the program executes the trade independently, removing the need for human intervention. This ability is crucial in volatile markets, where quick execution and precision can significantly affect results.
4) Customizability: Most trading bots provide customizable options, allowing users to set parameters like their risk tolerance and preferred trading strategies.

Advantages of trading bots

Speed and efficiency: trading bots can react to even the smallest changes with lightning-fast actions. This far outcompetes manual trading, as humans are usually not observing the market the entire time it is open.

Lack of emotion: trading bots do not go through uncertainty, fear, greed, or hesitation. Many manual traders go through a “Should I? Shouldn’t I?” moment when choosing to buy or sell stocks, and this often leads to loss due to missed trades, rapidly declining stock value, or simply stocks that have been held for far too long in the hopes of increased value. Bots have no such qualms, and can easily make trades based purely on mathematical formulas and data analysis.

24/7 market access: Unlike humans, bots don’t need to sleep. Even when the stock market is closed, they can monitor after hours prices and indices to determine the best course of action as soon as the market opens.

Backtesting and data analysis: Historical data can be used to simulate trading strategy performance, setting a precedent and helping evaluate the potential of said strategy on future trades. This is akin to a practice simulation.

]Disadvantages of trading bots

Frequent monitoring required: Despite their benefits, trading bots cannot completely eliminate user involvement. Regular evaluation of the bot’s performance, along with adjustments to respond to shifting market conditions, is crucial for limiting losses. Implemented risk-control

measures, such as stop-loss orders or portfolio diversification, help reduce potential setbacks. Users may also stay alert to security concerns like vulnerable APIs and unauthorized access.

Volatility challenges: Trading bots are susceptible to losses when market conditions change abruptly. A bot designed to follow a particular trend may have difficulty adapting to new circumstances like breaking news or sudden crashes, which can lead to financial setbacks.

Strategy dependence: A trading bot's effectiveness is tied directly to the quality of its coding and its strategies. Since bots cannot foresee future market movements, they depend on historical data and preset rules. Furthermore, the bot's ability to operate efficiently in volatile markets largely hinges on the specific approach it employs, such as arbitrage or scalping.

Purpose of research

This study was conducted to determine the effectiveness of five specific trading bots compared to traditional trading strategies like Mean Variance and DJIA. After the initial run of all notebooks, modifications were implemented on the bots in order to observe the impact on performance. These modifications are detailed later.

Methodology

This paper tests trading bots featured in a FinRL GitHub repository. Financial reinforcement learning, or FinRL, is the first open-source framework for financial reinforcement learning developed for the AI4Finance Foundation, a nonprofit organization. FinRL has three layers: market environments, agents, and applications. In order to complete a trading task, an agent interacts with the environment to make sequential decisions. The repository features a train-test-trade pipeline. The "example" folder, the one primarily used in the research study, features trading data files, the bots themselves, and results files. Data is sourced from around fifteen different locations, such as YahooFinance, QuantConnect, and WRDS. Specifically, the trading bots in the Stock NeurlPS2018 series were employed.

The FinRL Github repository offers five different trading bot strategies as well as testing against two manual strategies. All of these work in distinct ways, some more similar to each other than others:

A2C (Advantage Actor Critic): A2C is named after its usage of reinforcement learning, with actors that choose actions and critics that evaluate said actions. A2C specifically is an advanced version of this approach. Instead of the critic simply informing the actor of its reward value, the critic evaluates the effectiveness of an action by comparing the actual reward to an expected reward. This strategy is fast to train and offers stable performance. Its biggest strength is its careful balance of exploration of outcomes and exploitation of opportunities (Simonini 2022).

DDPG (Deep Deterministic Policy Gradient): DDPG is designed for continuous action spaces. Practically, this means that instead of simple actions like buy, sell, or hold, the bot also decides *how much* to buy or sell. It combines Q-learning, a value-based algorithm, and policy gradient methods, a class of algorithms that directly learn the policy. Similar to A2C, DDPG uses

the actor neural network (reliant on the policy gradient methods) and the critic neural network (reliant on Q-learning). This strategy is good for fine-tuned portfolio allocation. Its biggest strength is its ability to make precise trade decisions (hence its use in continuous action spaces) (PythonProg 2023).

TD3 (Twin Delayed Deep Deterministic Policy Gradient): TD3 is a successor to DDPG. Although DDPG has been very popular, it can be unstable and heavily reliant on finding the correct hyperparameters for the current task. The first big difference is that TD3 employs two critics rather than one, and uses the smallest return value of the two. The second is that it delays updates to the actor to ensure better critic accuracy. The third is noise regularization, where random noise is added to the target so the bot prefers higher values for actions that are more robust. These changes made TD3 better at avoiding false opportunities and making smarter trades than DDPG (Byrne 2019).

PPO (Proximal Policy Optimization): PPO is an on-policy strategy, which means that it learns from the actions taken within its current policy rather than a different set of data. Like other strategies, it uses an actor-critic algorithm. However, it is different in that it prevents the bot from making drastic changes to its policy—methodology for decision making—by adding a trust region. This makes PPO very stable and reliable for training. However, its biggest strength is also its biggest weakness: PPO tends to be a little too cautious in volatile environments, the likes of which are often opportunity-rich. This can lead to lower returns (Santosh 2023).

SAC (Soft Actor-Critic): SAC, belonging to the family of Actor-Critic algorithms (similar to A2C), is an off-policy reinforcement learning algorithm tailored for continuous action spaces. SAC's main objective is to maximize reward and policy entropy (randomness, which promotes exploration). It thus achieves a near-perfect balance between exploitation and exploration, making it excellent for complex, unpredictable environments. However, a challenge of this algorithm is that it requires very careful tuning of parameters like target entropy and replay buffer size (Qui 2024).

Mean-Variance Analysis: Mean-Variance is the first of the unautomated strategies tested in this paper. Its goal is to either maximize return for a given risk or minimize risk for a given return. Mean-Variance has two main components: variance, a number representing how spread out the numbers are in a set, and expected return. These two components are used to calculate the weight of each asset and thus the portfolio's expected return, making this strategy very mathematically sound (Chen 2025).

DJIA (Dow Jones Industrial Average): DJIA is a stock market index that consists of 30 large-cap companies such as Apple, NVIDIA, and Microsoft. The strategy used as a method of comparison is the purchase of all 30 stocks in the DJIA and holding them for an extended period of time. This is a passive, low-cost manual strategy that relies on long-term growth. This strategy is weak, as it has no adaptation to market changes, misses out on active opportunities, and cannot protect against market downturns (Adkins 2024).

Testing Protocol: This series has three notebooks, or collections of code, that are used for the training and testing of all reinforcement-learning strategies. Notebook 1 is responsible for fetching trading data from Yahoo Finance through the APIs in the yfinance open-source library. The data frame is adjusted in a form that is convenient for further data processing. The code then checks for missing data and converts the data point into a state. Technical indicators such as the MACD (moving average convergence/divergence) and a turbulence index are added. The data is split and saved to a CSV file used by Notebook 2.

Notebook 2 builds a market environment for reinforcement learning in OpenAI Gym-style. This features an agent that is active in a world called the environment. This agent observes its current condition as a state and is allowed to do certain actions that modify the state. Simultaneously, the environment will have feedback called "rewards" that indicate the success of the agent's actions. The goal of the agent is to get as much cumulative reward as possible, a process that leads to an optimized model. This notebook first reads the .CSV data, constructs the environment, then trains the RL agents through the aforementioned reward-based method. Five algorithms are employed: A2C, DDPG, PPO, TD3, and SAC. Their use cases will be further outlined later. After training, the agents are saved and used in Notebook 3.

Notebook 3 uses the trained agents to conduct a backtest and compare with baselines such as Mean Variance Optimization and the DJIA index. Backtesting plays a key role in evaluating the performance of a trading strategy. This notebook uses an automated backtesting tool through the Quantopian PyFolio package. This tool also generates a returns plot for all trading strategies. This notebook was modified for the purpose of this research study to compute and return a table of the following important performance metrics: Total Return %, Sharpe Ratio, Volatility (%), Max Drawdown (%), Trend Slope, and R^2 score.

All training strategies output six distinct metrics (Nath 2024).

1) Total return (%): This is the percentage gain or loss on an investment over a given time frame. It is different from price return as it accounts for both price appreciation (stock fluctuation) and income (dividends, interest, etc.).
2) Sharpe ratio: This measures the return per unit of risk. A higher Sharpe ratio is preferable as it shows high return for the amount of risk taken.
3) Volatility (%): This measures the fluctuation of returns over time (the standard deviation of returns). A lower volatility is often preferred because it indicates higher stability, while a high volatility indicates potential for both large losses and large gains.
4) Max drawdown (%): This measures the largest peak-to-trough decline during a given time frame. A lower max drawdown is preferable as it indicates risk-averseness.
5) Trend slope: This indicates the progression of a best-fit line, found by applying linear regression to price data. A positive slope indicates rising prices, while a negative slope indicates the opposite. A higher trend slope is generally preferred.

6) R^2: This measures how much of an asset's movements are explained by movements in a given benchmark (like the S&P 500). An R^2 close to 1 indicates that the asset is highly correlated with the benchmark, which is good for broad market exposure but less useful for diversification. A low R^2 indicates that the asset moves differently from the market. For the purposes of this paper and the time period of the test data, a high R^2 is preferred due to its strong correlation with the benchmark.

These metrics were tested for all strategies by running the three FinRL notebooks from scratch. This resulted in freshly generated test and training data. The last notebook was slightly modified in order to produce a clean output of the resulting metrics.

The initial backtest was conducted with the default values for variables such as volatility gate, HMax, and selling cost. After this test, all these values were modified separately in order to observe the changes in the results.

The first modification was an attempt to alter trading frequency. Oftentimes, frequent trading can reduce performance if the bot overreacts to small market changes. Alternatively, infrequent trading can also reduce performance due to lowered reactions to important market changes. This was tested by modifying a variable controlling transaction cost (originally 0.1%) to values of 0.0%, 0.2%, and 0.3%. A higher transaction cost leads to lower frequency of trade.

The second modification was to avoid or increase trading during high volatility periods. Changing the volatility gate is a way to control the level of volatility in which the bots are allowed to trade. Lowered volatility will require more stable conditions, while higher volatility allows for unstable but sometimes more profitable conditions. This was tested by modifying a variable controlling volatility gate (originally 70) to 30, 40, and 50.

The last modification was to reduce the HMax. HMax reduction means that the bot needs stronger signals to make a trade. This is often a way to avoid smaller and sometimes less profitable trades. This was tested by modifying HMax (originally 100) to 10, 50, and 90.

All these tests generated the same graph and table. The values were then analyzed through a comparison to the original, unmodified metrics.

FinRL Bot Analysis

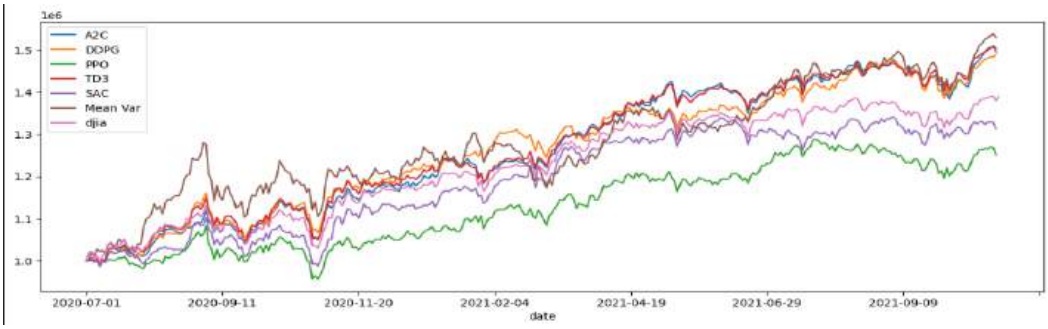


Figure 1.1: Graph detailing base performance of all FinRL strategies

Full Strategy Evaluation:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	50.27	2.37	13.35	-7.87	1528.52	0.954
1	AI Bot (DDPG)	48.87	2.15	14.48	-9.01	1418.71	0.956
2	AI Bot (PPO)	25.11	1.24	14.47	-11.62	945.98	0.910
3	AI Bot (TD3)	49.51	2.21	14.17	-8.81	1494.65	0.957
4	AI Bot (SAC)	31.23	1.47	14.73	-10.32	1142.36	0.904
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 1.2: Metric comparison of all unmodified FinRL Strategies

All three notebooks were run so the training data was fresh. In this evaluation, the mean-variance strategy achieved the highest return at 52.63%, outperforming the closest bot (A2C) by roughly 2%. However, the mean-variance strategy had a Sharpe ratio of 1.71, significantly lower than A2C's 2.37. The higher Sharpe ratio indicates better performance. Additionally, mean-variance recorded nearly 20% volatility, much higher than the 14% average of other strategies, while A2C had the lowest volatility at 13.35%, showing the stability of A2C.

Similarly, mean-variance had the highest (most negative) max drawdown, while A2C had the lowest. The trend slope of mean-variance ranked around the middle of all strategies, and its R^2 score was the lowest. In contrast, A2C had the highest trend slope and a strong R^2 score.

The worst-performing strategy was the PPO AI bot, with a return of only 25.11% and poor results across all other metrics. All other strategies performed mediocrely with few standout performances. DDPG and TD3 had similar performances, just behind mean-variance and A2C. SAC's performance lagged slightly behind DDPG and TD3 but remained stronger than PPO. DJIA performed worse than A2C, DDPG, and TD3, but better than PPO and SAC.

Overall, while mean-variance had the highest total return, this simulation is that of a stable market. It performed poorly in nearly all other metrics, meaning it cannot be relied on for stability, growth, and optimization in more tumultuous conditions.Just behind it in total return was the A2C AI bot, only 2% lower but among the best in every other category. Therefore, A2C can be considered the optimal strategy of those tested, balancing risk and return remarkably well.

FinRL Bot Modifications: Altering Trading Frequency

Trading bots can be affected by various environmental and operational factors. One such is trading frequency: the amount of trades a bot is permitted to make. This can be controlled by the transaction cost of each trade. Each bot has a "budget," thus lowered transaction costs encourage higher frequency of trades while higher transaction costs do the opposite. The question to examine is: How does trading frequency affect bot performance

Frequent trading can reduce performance if the bot overreacts to small market changes, but infrequent trading can also reduce performance due to a slow reaction to important market

changes. Therefore, it can be hypothesized that there is a balance, where both excessive and insufficient trading frequency may reduce effectiveness.

In order to test this, transaction cost was adjusted in Notebook 3 of the FinRL repository. value of 0.1% was compared to modifications at 0.0%, 0.2%, and 0.3%. Lower costs lead to increased frequency in trading, while higher costs have the opposite effect.
Results for 0.0% (increase in frequency):

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	51.76	2.39	13.54	-7.98	1564.10	0.957
1	AI Bot (DDPG)	49.16	2.16	14.49	-9.02	1422.53	0.956
2	AI Bot (PPO)	25.23	1.24	14.48	-11.61	948.19	0.911
3	AI Bot (TD3)	49.69	2.22	14.17	-8.81	1496.67	0.957
4	AI Bot (SAC)	31.39	1.47	14.72	-10.31	1144.00	0.904
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 2.1: Metric comparison of FinRL strategies with 0.0% transaction costs

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	+1.49	+0.02	+0.19	-0.11	+35.58	+0.003
DDPG	+0.29	+0.01	+0.01	-0.01	+3.83	0
PPO	+0.12	0	+0.01	+0.01	+2.21	+0.01
TD3	+0.18	+0.01	0	0	+2.02	0
SAC	+0.16	0	+0.01	-0.01	+1.64	0
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 2.2: Analysis of metric changes from unmodified to modified transaction cost (+- represents changes between the original parameter value and the newly set value)

With the decrease of transaction cost from 0.1% to 0.0%, the bot most impacted was A2C (Figure 2.2). Its total return increased by 1.49%, and all other metrics improved (most notably the trend slope, which rose by 35.58), except for volatility. As mentioned previously, lower volatility is preferred among trading bots; in this case, volatility increased slightly by 0.19%. However, this change is minimal compared to the significant improvements in other metrics, particularly total return and trend slope.

Other bots showed similar effects, but to a lesser extent. The remaining four bots averaged a 0.1875% increase in total return, a 0.005 increase in Sharpe ratio, and a negligible

rise in volatility, with no changes in max drawdown or R^2 score. Naturally, Mean-Variance and DJIA saw no changes, as they are not bot-driven strategies.

Overall, these results suggest that a reduction in transaction cost is most beneficial for A2C, and marginally improves all other strategies as well.

Results for 0.2%:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	50.09	2.36	13.36	-7.87	1526.45	0.954
1	AI Bot (DDPG)	48.63	2.14	14.48	-9.03	1414.47	0.956
2	AI Bot (PPO)	24.90	1.23	14.46	-11.62	941.29	0.910
3	AI Bot (TD3)	49.32	2.21	14.17	-8.81	1492.55	0.957
4	AI Bot (SAC)	31.08	1.46	14.74	-10.33	1140.66	0.903
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 2.4: Metric comparison of FinRL strategies with 0.2% transaction costs

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	-0.18	-0.01	+0.01	0	-2.07	0
DDPG	-0.24	-0.01	0	-0.02	-4.24	0
PPO	-0.21	-0.01	-0.01	0	-4.69	0
TD3	-0.19	0	0	0	-2.10	0
SAC	-0.15	-0.01	+0.01	-0.01	-1.70	-0.001
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 2.5: Analysis of metric changes from unmodified to modified transaction cost (+- represents changes between the original parameter value and the newly set value)

With the increase in transaction costs from 0.1% to 0.2%, all bots faced a decline in total return by an average of 0.194% (Figure 2.5). In conjunction with the marginal decreases in Sharpe ratio and trend slope, this indicates that this algorithmic change results in a negative impact, however slight.

Volatility showed negligible average change, with some bots staying stagnant while others varied by only ±0.01. Max drawdown followed a similar pattern, with three bots unchanged and two showing fractional decreases.

Although increasing the transaction costs by 0.1% had a negligible negative impact, it can be concluded that this strategy will not result in improved bot performance.

Results for 0.3%

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	49.79	2.35	13.35	-7.87	1521.72	0.954
1	AI Bot (DDPG)	48.43	2.13	14.50	-9.05	1410.22	0.957
2	AI Bot (PPO)	24.83	1.23	14.45	-11.62	939.35	0.910
3	AI Bot (TD3)	49.14	2.20	14.17	-8.81	1490.59	0.957
4	AI Bot (SAC)	30.92	1.45	14.74	-10.34	1138.80	0.903
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 2.7: Metric comparison of FinRL strategies with 0.3% transaction costs

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	-0.48	-0.02	0	0	-3.80	0
DDPG	-0.44	-0.02	+0.02	+0.02	-8.49	+0.001
PPO	-0.28	-0.01	-0.02	-0.02	-6.63	0
TD3	-0.37	-0.01	0	0	-4.06	0
SAC	-0.31	-0.02	-0.02	+0.01	-3.56	-0.001
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 2.8: Analysis of metric changes from unmodified to modified transaction cost (+- represents changes between the original parameter value and the newly set value)

With the increase in transaction costs from 0.1% to 0.3%, all bots faced a decline in total return (Figure 2.8). A2C and DDPG were impacted the most, with an average drop of 0.46%, while PPO, TD3, and SAC averaged a 0.32% drop. DDPG suffered the most in terms of trend slope (by 8.49), followed closely by PPO. A2C, TD3, and SAC averaged a drop of 3.81 in trend slope. All bots faced a drop in Sharpe ratio of either 0.01 or 0.02 percent.

Volatility and max drawdown remained largely unchanged, varying by no more than ±0.02%. R^2 score remained largely the same as well. These impacts, however small in magnitude, show that increased transaction costs have the opposite of the intended effect. Rather than significantly improving bot performance, this adjustment slightly worsens it.

Overall, these tests show that reducing transaction costs benefits trading bot performance, while increasing them leads to small declines. These changes most significantly impacted A2C and DDPG. This data suggests that frequent trading can be advantageous when costs are low,

since it allows bots to exploit more short-term opportunities without the penalties incurred by high fees. Conversely, raising transaction costs diminishes returns because it reduces profits from each trade.

FinRL Bot Modifications: Avoiding Trading During High Volatility Periods

Volatility gate, also known as the turbulence threshold, determines the level of market volatility in which the bots are permitted to operate. Therefore, changes to this variable directly test the impact of market stability on trading results.

The expectation is that a lowered volatility gate (or turbulence threshold) will force bots to trade in calmer, more stable markets. This could improve performance by avoiding trade in chaotic conditions. Conversely, a higher volatility gate, although associated with higher risk, can allow for more fruitful trades.

In order to test this, a variable called turbulence threshold was adjusted in the FinRL repository. The initial turbulence threshold was set at 70. Results at 30, 40, and 50 follow, where lower values indicate a more strict trading setup. Higher levels of 60, 80, 90, and 100 were also tested, but no changes were observed. This likely means that volatility never goes above 50 in the dataset on which the bots are operating, thus any higher gates will have no impact.

Results for 30:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	39.04	2.15	11.91	-6.35	1154.80	0.945
1	AI Bot (DDPG)	38.09	1.97	12.78	-7.18	1034.93	0.943
2	AI Bot (PPO)	17.51	1.16	11.05	-6.82	628.87	0.850
3	AI Bot (TD3)	38.47	2.04	12.42	-7.32	1146.68	0.952
4	AI Bot (SAC)	23.64	1.30	12.96	-6.60	875.14	0.916
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 3.1: Metric comparison of FinRL strategies with volatility gate of 30

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	-11.23	-0.22	-1.44	+1.52	-373.72	-0.009
DDPG	-10.78	-0.18	-1.7	+1.83	-383.78	-0.013
PPO	-7.60	-0.08	-3.42	+4.80	-317.11	-0.060
TD3	-11.04	-0.17	-1.75	+1.49	-347.97	-0.005
SAC	-7.59	-0.17	-1.77	+3.72	-257.22	+0.012
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 3.2: Analysis of metric changes from unmodified to modified volatility gate

With the decrease of the volatility gate (also known as turbulence threshold) from 70 to 30, all bots experienced significant negative impacts (Figure 3.2). This was felt most heavily by A2C, with a decline in return percent by 11.23, a drop in trend slope by 373.72, and a decrease in Sharpe ratio of 0.22. DDPG and TD3 averaged a decline in total return percent of 10.91. PPO and SAC averaged a decline in total return percent of 7.60. Max drawdown and the R^2 score showed negligible changes (with the exception of PPO, with an R^2 drop of 0.060 and max drawdown increase of 3.42). The only positive effect from this change was a reduction in volatility, which again was most notable in PPO. These impacts show that a decreased volatility gate results in reduced returns and weaker performance across all metrics.

Results for 40:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	46.28	2.25	13.14	-7.57	1370.38	0.951
1	AI Bot (DDPG)	46.21	2.07	14.37	-9.01	1322.50	0.959
2	AI Bot (PPO)	16.47	0.90	13.76	-11.66	623.33	0.745
3	AI Bot (TD3)	45.22	2.10	13.90	-8.81	1354.09	0.954
4	AI Bot (SAC)	28.57	1.38	14.51	-10.18	1035.34	0.900
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 3.4: Metric comparison of FinRL strategies with volatility gate of 40

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	-3.99	-0.12	-0.21	-0.30	-158.14	-0.003
DDPG	-2.66	-0.08	-0.11	0	-96.21	+0.003
PPO	-8.64	-0.34	-0.71	-0.04	-322.65	-0.165
TD3	-4.29	-0.11	-0.24	0	-140.56	-0.003
SAC	-2.66	-0.09	-0.22	+0.14	-107.02	-0.004
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 3.5: Analysis of metric changes from unmodified to modified volatility gate

With the decrease of the volatility gate (also known as turbulence threshold) from 70 to 40, all bots experienced significant negative impacts (Figure 3.5). This was felt most heavily by PPO, with a decline in return percent by 8.64, a drop in trend slope by 322.65, and a drop in Sharpe ratio of 0.34. The remaining bots averaged a decline in total return percent of 3.4, drop in Sharpe ratio of 0.1, and a decrease in trend slope of 125.48.

Max drawdown and the R^2 score showed negligible changes (with the exception of PPO, with an R^2 drop of 0.1650). The only positive effect from this change was a reduction in volatility, which again was most notable in PPO. These impacts show that a decreased volatility gate results in reduced returns and weaker performance across all metrics.
Results for 50:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	50.27	2.37	13.35	-7.87	1528.52	0.954
1	AI Bot (DDPG)	48.87	2.15	14.48	-9.01	1418.71	0.956
2	AI Bot (PPO)	25.11	1.24	14.47	-11.62	945.98	0.910
3	AI Bot (TD3)	49.51	2.21	14.17	-8.81	1494.65	0.957
4	AI Bot (SAC)	31.23	1.47	14.73	-10.32	1142.36	0.904
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 3.7: Metric comparison of FinRL strategies with volatility gate of 50

Changing the volatility gate from 70 to 50 does not alter the results in the slightest (Figure 3.7). This leads to the conclusion that volatility never goes above 50 (between 50 and 70) in the dataset on which the bots are operating, thus any higher gates will have no impact. This is a limitation of this specific study and dataset. Employing other datasets with higher volatility may impact the results differently.

Overall, lowering the volatility gate from 70 weakened trading bot performance, with sharper declines the lower the threshold. This happened because a stricter volatility filter prevented the bots from participating in more lucrative trades, even when market conditions weren't extremely adverse. The only benefit in the metrics was the expected reduction in volatility, but this came at the cost of significantly lower returns, weaker trend slopes, and reduced Sharpe ratios across all bots. Since the dataset's volatility never exceeded 50, raising the threshold showed no evidence of change. In conclusion, there are many negative impacts at lower volatility gate levels due to aggressive restrictions, while higher thresholds show no effect because the market conditions never reached those extremes.

FinRL Bot Modifications: Changing HMax

In the FinRL repository, HMax determines the strength of signals that are required before a bot is permitted to trade. A lower HMax forces bots to wait for stronger confirmation before trading, thus lowering it tests the impact of stricter entry requirements.

If HMax is reduced moderately, it may filter out trades, improving performance by focusing only on high-confidence opportunities. However, if it is reduced too drastically, it could prevent bots from reacting quickly to market changes, leading to missed opportunities and performance declines.

In order to test this, HMax was varied across several different trials. The initial HMax was set at 100. Levels of 10, 50, and 90 were tested. 100 describes the maximum amount for HMax, thus no higher levels were tested.
Results for 10:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	40.30	2.08	12.65	-8.61	1321.31	0.949
1	AI Bot (DDPG)	41.88	2.02	13.55	-7.96	1295.56	0.960
2	AI Bot (PPO)	11.04	1.19	6.85	-6.42	401.38	0.859
3	AI Bot (TD3)	41.25	1.98	13.64	-8.98	1343.32	0.957
4	AI Bot (SAC)	27.23	1.41	13.54	-9.19	1046.80	0.909
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 4.1: Metric comparison of FinRL strategies with HMax of 10

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	-9.97	-0.29	-0.7	-0.74	-207.21	-0.005
DDPG	-6.99	-0.13	-0.93	+1.05	-123.15	+0.004
PPO	-14.07	-0.05	-7.62	+5.2	-544.6	-0.051
TD3	-8.26	-0.23	-0.53	-0.17	-151.33	0
SAC	-4	-0.06	-1.19	+1.13	-95.57	+0.005
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 4.2: Analysis of metric changes from unmodified to modified HMax

With the decrease of HMax from 100 to 10, all bots experienced significant negative impacts (Figure 4.2). This impacted PPO the most, with a decline in return percent of 14.07, a decline in trend slope of 544.6, and a decline in R^2 of 0.051. A2C and TD3 were just under that with declines in return percent of around 9, with DDPG and SAC trailing further behind. Sharpe ratio declined significantly for A2C and TD3, but that difference was less notable for all other bots. Max drawdown and R^2 had negligible changes, while trend slope for all bots other than PPO declined by an average of 144.32. The only positive impact from this change was a decrease in volatility (most notable in PPO), but the negatives far outweigh that lone positive.
Results for 50:

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	52.40	2.41	13.60	-8.14	1629.60	0.960
1	AI Bot (DDPG)	48.53	2.15	14.36	-8.95	1417.70	0.959
2	AI Bot (PPO)	23.82	1.26	13.50	-9.91	906.24	0.906
3	AI Bot (TD3)	48.91	2.20	14.08	-8.97	1484.66	0.959
4	AI Bot (SAC)	30.71	1.48	14.40	-10.05	1128.93	0.905
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 4.4: Metric comparison of FinRL strategies with HMax of 50

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	+2.13	+0.04	+0.25	-0.27	+101.08	+0.006
DDPG	-0.34	0	-0.12	+0.06	-1.01	+0.003
PPO	-1.29	+0.02	-14.47	+1.71	-39.74	-0.004
TD3	-0.6	-0.01	-0.09	-0.16	-9.99	+0.002
SAC	-0.52	+0.01	-0.33	+0.27	-13.43	+0.001
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 4.5: Analysis of metric changes from unmodified to modified HMax

With the decrease of HMax from 100 to 50, most bots experienced a very small negative impact (Figure 4.5). In the case of A2C, there was actually a small positive impact, as all metrics moved in a desirable direction—though volatility increased slightly, which is expected with higher returns. PPO suffered the most, relatively speaking, with a return percent decline of 1.29 compared to the other bots' (excluding A2C) average of 0.49. However, PPO also reaped the highest benefit in terms of volatility drop and max drawdown. Aside from that, there were negligible changes in R^2, max drawdown, Sharpe ratio, and trend slope. Thus it can be concluded that changing the HMax from 100 to 50 does not have a significant impact.

Results for 90

	Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
0	AI Bot (A2C)	53.06	2.42	13.69	-8.10	1619.57	0.957
1	AI Bot (DDPG)	49.16	2.16	14.45	-8.90	1419.75	0.957
2	AI Bot (PPO)	24.69	1.23	14.38	-11.68	930.15	0.908
3	AI Bot (TD3)	48.91	2.19	14.18	-8.97	1465.51	0.958
4	AI Bot (SAC)	30.22	1.44	14.62	-10.34	1097.82	0.901
5	Mean-Variance	52.63	1.71	19.76	-13.77	1250.06	0.893
6	DJIA (Buy & Hold)	38.84	1.86	13.76	-8.93	1173.05	0.926

Figure 4.7: Metric comparison of FinRL strategies with HMax of 90

Strategy	Total Return (%)	Sharpe Ratio	Volatility (%)	Max Drawdown (%)	Trend Slope	R^2 Score
A2C	+2.79	+0.05	+0.34	-0.23	+91.05	+0.003
DDPG	+0.29	+0.01	-0.03	+0.11	+1.04	+0.001
PPO	-0.42	-0.01	-0.09	-0.06	-15.83	-0.002
TD3	-0.6	-0.02	+0.01	-0.16	-29.14	+0.001
SAC	-1.01	-0.03	-0.11	-0.02	-44.54	-0.003
Mean-Variance	0	0	0	0	0	0
DJIA	0	0	0	0	0	0

Figure 4.8: Analysis of metric changes from unmodified to modified HMax

With the decrease of HMax from 100 to 90, two of the bots faced a marginal positive impact while the other three faced a marginal negative one (Figure 4.8). Namely, A2C and DDPG's return percentages rose, although A2C had a relatively higher increase. The reductions in the remaining three bots' return percent were an average of 0.68. Small negative changes were observed in the other metrics, especially for A2C where volatility increased by 0.34 and max drawdown decreased by 0.23. The remaining metrics had negligible changes across the board. These results lead to the conclusion that this slight reduction in HMax can lead to either small boosts in performance, as seen with A2C and DDPG, or other negligible changes.

Overall, the HMax tests demonstrate that requiring stronger trading signals generally harms both performance when taken to an extreme, but has minimal or mixed effects with moderate adjustment. Reducing HMax from 100 to 10 significantly limited trade opportunities, which caused substantial drops in returns, trend slope, and Sharpe ratio across all bots. This shows that overly restrictive signal thresholds prevent these bots from capitalizing on profitable trades. However, reducing HMax to 50 or 90 led to only minor impacts, with some bots (notably

A2C) experiencing small gains while others declined slightly. This suggests that small reductions of HMax does not meaningfully improve performance. In short, while extremely low HMax constrains both effectiveness, moderate reductions result in negligible or inconsistent performance changes.

Conclusion

In this study, the open-source FinRL repository was used to run multiple trading bots, each based on different training strategies, allowing for a systematic evaluation of their effectiveness relative to both index-based benchmarks and math-driven strategies. Of the trading bots–A2C, DDPG, PPO, TD3, and SAC–A2C performed the best, with a return of 50.27%. A2C was outperformed slightly by Mean-Variance, an equation-driven strategy often used by professionals in the finance industry. However, considering performance in all other tested metrics, it was determined that the A2C bot was the most optimal trading strategy.

The bot modifications that followed had mixed results. The first involved altering trading frequency through the change of transaction cost. When trading frequency was increased (transaction costs decreased to 0.0%), marginal positive changes were observed in all bots, with the highest difference in A2C. Contrastingly, decreasing trading frequency led to decreases in performance, although negligible.

The second modification involved changing the volatility gate. When reduced, there were very substantial declines in performance across all bots and metrics. Predictably, volatility was the only metric that showed an improvement. However, an increase in volatility gate could not be tested. The original level was 70, but results from 50 and all the way to 100 at intervals of 10 were identical to the original, leading to the conclusion that volatility never rises above 50.

The last modification involved reducing HMax, which meant trading bots needed stronger signals to make a trade. Based on results at 10, 50, and 90, it was found that reducing this metric in moderation may result in marginal improvements or stagnancy, while sharp reductions (at 10) resulted in significant performance declines. The improvements at 50 and 90 (affecting A2C the most) were close to insignificant, and at times inconsistent.

Overall, A2C had the strongest showing (compared to other bots) throughout all modifications, as well as the most consistent performance. Additionally, small parameter tweaks (transaction costs, HMax) had limited impact compared to more foundational elements such as volatility gate, which suggests that some model parameters are more influential than others. This high sensitivity to volatility threshold indicates that market stability is a significant factor in bot performance. This mimics real-world trading, where volatility can significantly diminish any trading strategies.

While the technical implications are important, this research also opens opportunities for future expansion:

1) Expand datasets and market conditions: Tests can be done across different financial environments, such as bull/bear markets, recessions, or extreme events

(e.g., COVID-19 or the 2008 crisis). Different types of assets can be included as well, such as cryptocurrencies, commodities, and forex. These tests would help determine whether bot resilience holds universally, or if it is tied to a particular dataset.

2) Longer simulation periods: These results cover only a one-year period. Multi-year backtesting can show whether these bots have the capability to adapt to long-term trends, or if they degrade over time.
3) Hybrid models: Exploration can be done regarding blended A2C with Mean-Variance, the two best performing overall strategies. In this case, the AI can adapt to market changes, while the math-driven model can provide stability. This may provide a "best of both worlds" scenario.
4) Adaptation to bull/bear markets: This is a bot modification that was not possible in the scope of this project. It would involve using indicators like 200-day averages, volatility, and other economic data to deduce whether the market is a bull or bear. Once the bot figures out the market state, it can adjust purchases such that more buying is done in bull (upward-trending) markets and less is done in bear (downward-trending) markets.

Research on this topic can also lead to many social impact applications:

1) Financial inclusion tools: Simplified versions of trading bots could be offered to small investors in underserved communities. This would give them increased access to strategies that are usually reserved for institutional traders.
2) Cooperative fund management: AI bots could help manage small-scale community/cooperative funds, making lending and saving more efficient for low-income groups.
3) Educational platforms: the same bots can be used in gamified platforms to teach students about finance, trading, and AI.

Overall, this research highlights that AI-driven trading bots can serve as powerful tools in finance when carefully designed, but their real-world effectiveness depends on adapting to market conditions and prioritizing the most influential model parameters.

Works Cited

AI4Finance Foundation. FinRL-Tutorials. GitHub, 2023, https://github.com/AI4Finance-Foundation/FinRL-Tutorials

"Trading Bot." Bitpanda Academy, Bitpanda, https://www.bitpanda.com/academy/en/lessons/trading-bot

"DDPG (Deep Deterministic Policy Gradient): A Beginner's Guide." PythonProg, https://pythonprog.com/ddpg-deep-deterministic-policy-gradient/

Santhosh, Sthanikam. "Reinforcement Learning (Part-8): Proximal Policy Optimization (PPO) for Trading Environment (TensorFlow)." Medium, 4 Mar. 2023, https://medium.com/@sthanikamsanthosh1994/reinforcement-learning-part-8-proximal-policy-optimization-ppo-for-trading-9f1c3431f27d

Byrne, Donal. "TD3: Learning To Run With AI." Medium, 5 May 2019, https://medium.com/data-science/td3-learning-to-run-with-ai-40dfc512f93

Qiu, Xuanyun. "Soft Actor-Critic Algorithm | From Theory to Practice." Medium, 5 Dec. 2019, https://medium.com/intro-to-artificial-intelligence/soft-actor-critic-reinforcement-learning-algorithm-1934a2c3087f

"Mean-Variance Analysis: Definition, Example, and Calculation." Investopedia, https://www.investopedia.com/terms/m/meanvariance-analysis.asp

"Strategies for Investing in the Dow Jones Industrial Average." Investopedia, https://www.investopedia.com/articles/financial-theory/10/introduction-to-the-dow.asp

"How Investment Risk Is Quantified." Investopedia, https://www.investopedia.com/articles/investing/032415/how-investment-risk-quantified.asp

"Slope Performance Trend." ChartSchool, StockCharts.com, https://school.stockcharts.com/doku.php?id=technical_indicators:slope_performance_trend

Hugging Face. "Advantage Actor Critic (A2C)." Hugging Face, 22 July 2022, https://huggingface.co/blog/deep-rl-a2c

Appendix

Following are graphs for each bot modification performance.

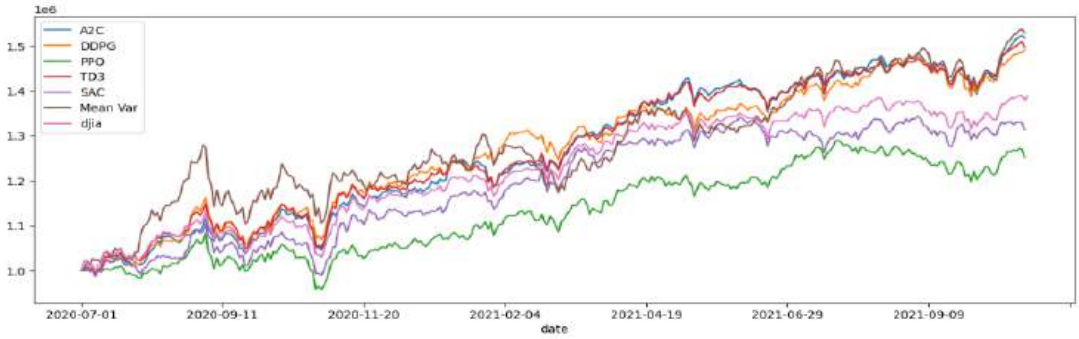


Figure 2.3: Results for 0.0% transaction cost

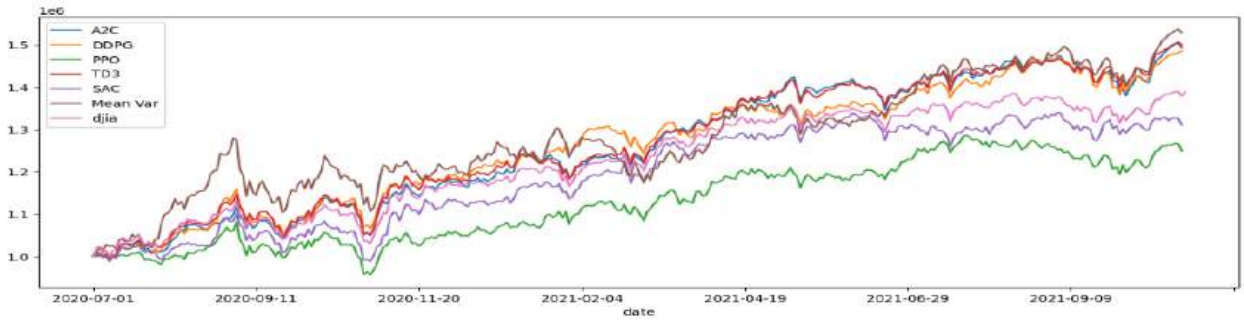


Figure 2.6: Results for 0.2% transaction cost

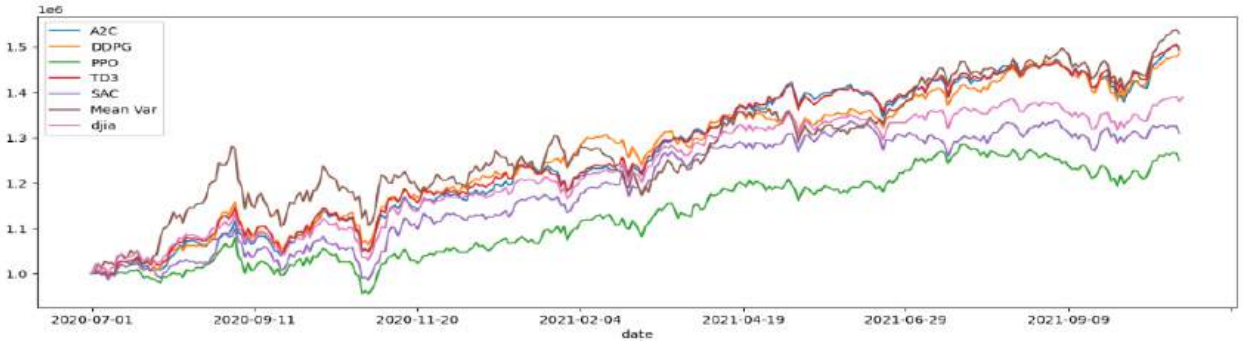


Figure 2.9: Results for 0.3% transaction cost

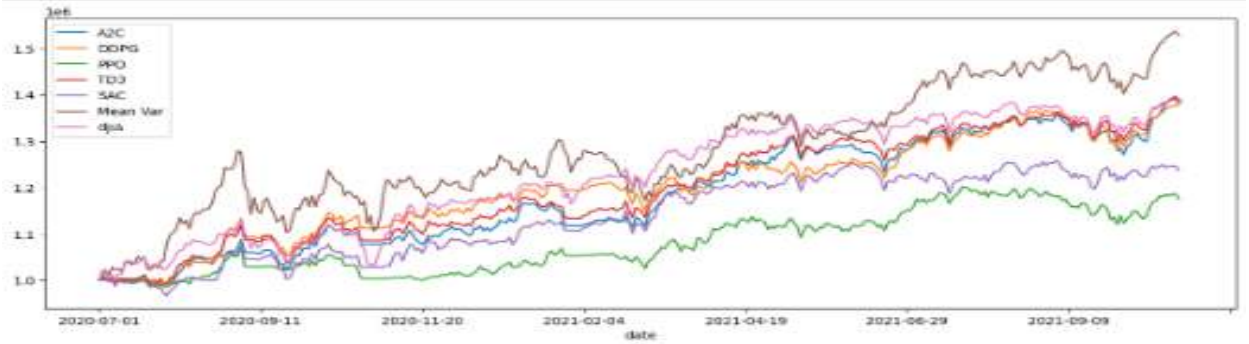


Figure 3.3: Results for turbulence threshold of 30

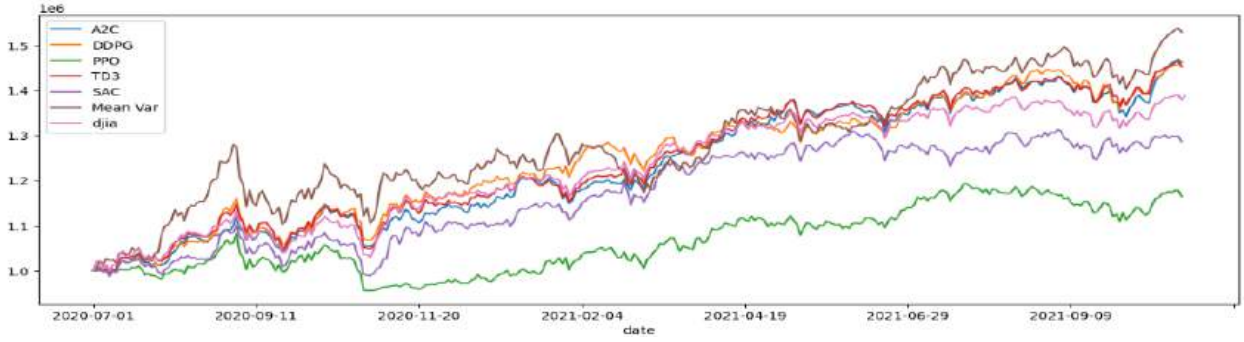


Figure 3.6: Results for turbulence threshold of 40

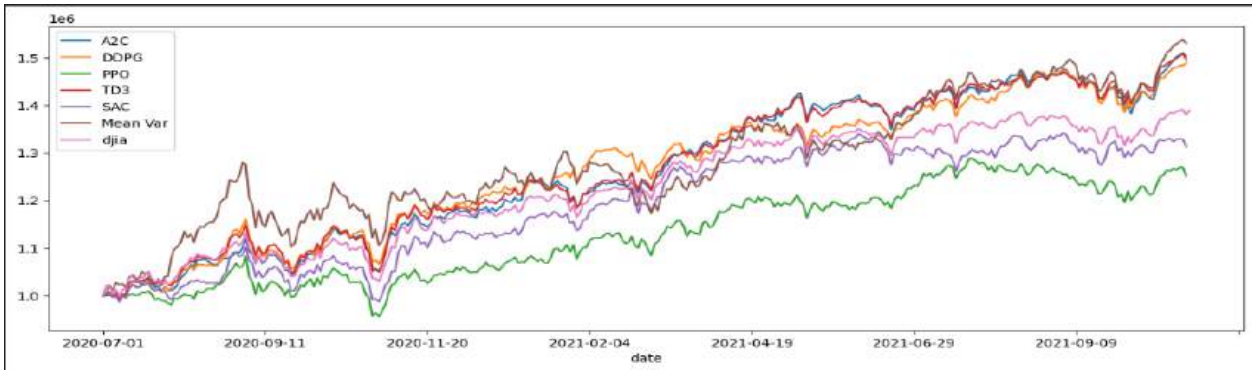


Figure 3.8: Results for turbulence threshold of 50

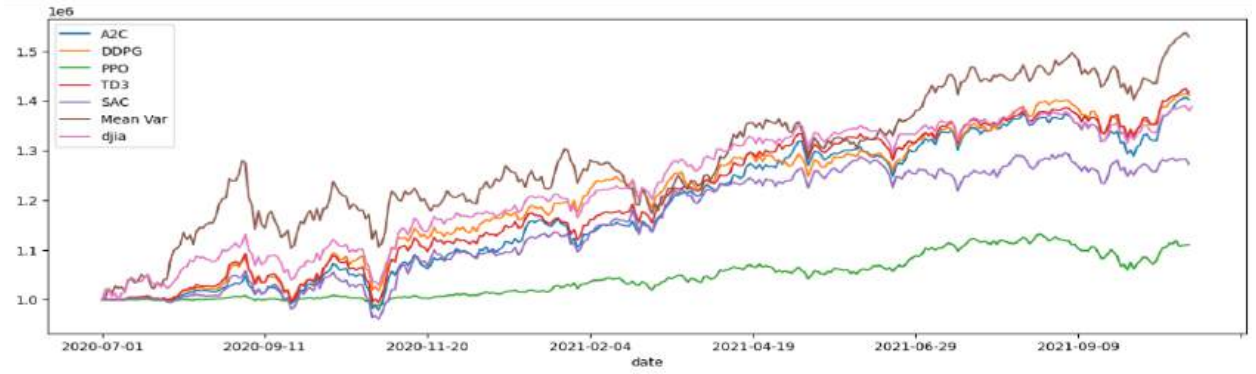


Figure 4.3: Results of HMax level of 10

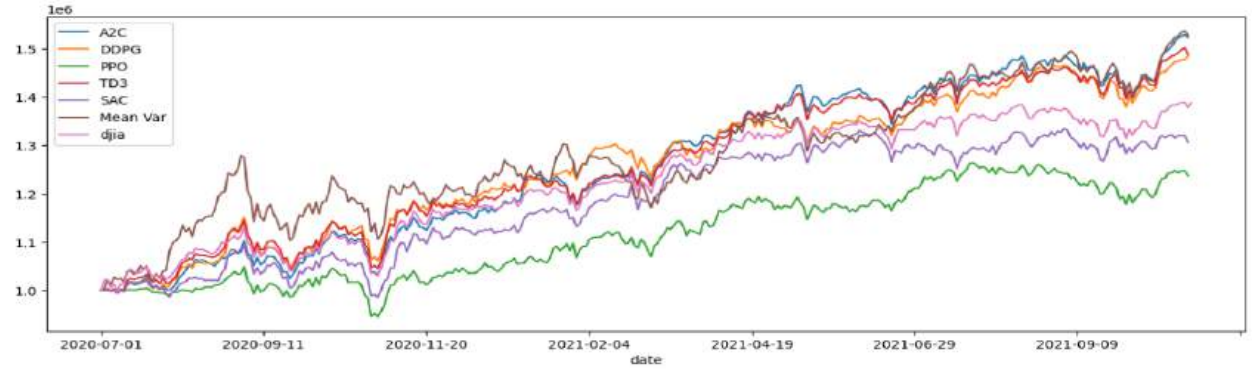


Figure 4.6: Results of HMax level of 50

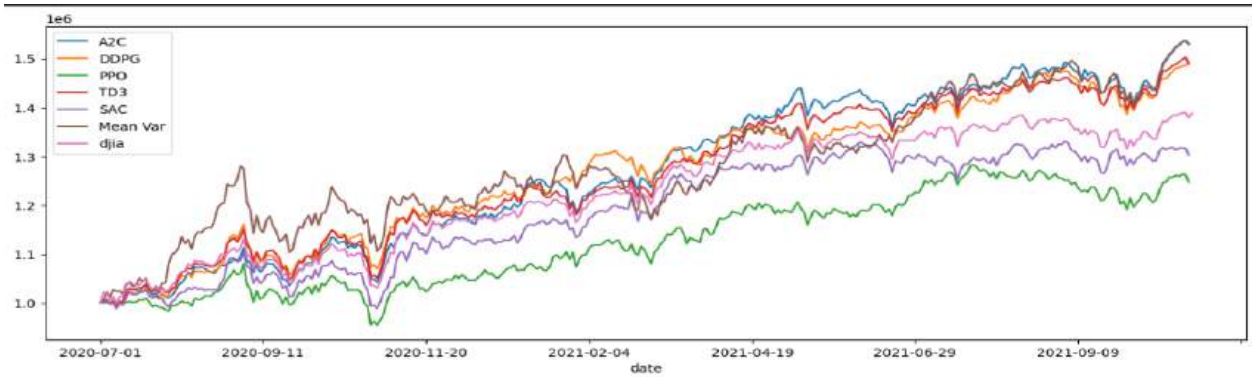


Figure 4.9: Results of HMax level of 90

Democratic Backsliding By Malin Mehfar

Abstract

To question whether democracy is in crisis is an idea that must be examined more carefully in this day and age. In the 20th century the world witnessed two periods of democratic crises during WWII and during the Cold War. These two events are often referenced when attempting to understand our current democratic status. We currently face a period of democratic backslidding, and based on past events it is likely to lead to a democratic crisis in the United States due to expansion of military power through the executive branch and the bypassing of congressional approval.

Introduction

Whether democracy is in crisis is a question that must be examined more carefully in this day and age– not only to prevent future crises, but also to understand the future potential of our current democratic backsliding. Democratic backsliding is the gradual process by which elected officials within a government erode democratic processes and norms (Bermeo). However, to truly examine this idea, we must first understand democracy. Democracy is defined as a government that places the power in the people's hands to create free and fair elections, political equality, and the protection of human rights (Shapiro). Athenian democracy, established from the 4th to 5th century BCE, was the first documented democracy that gave the people the power to vote for their leaders (Cartwright). This created a precedent for democracies like Iceland, San Marino, and later the USA. However, in recent years, the people have felt their constitutional rights being tested, while democratic countries increasingly drift towards autocracy.

But what we really need to take into account at this point is the wording of the term ‘crisis’: not only what it means, but also whether it applies to our current situation. The word crisis is defined as: a turning point, or a decisive moment in a situation when other solutions seem to fall short, causing danger or distress (“Crisis.”). Our world might be at that tipping point; however, to really understand our current democratic stance we must first examine past democratic crises.

One of the first European democracies to fall during WWII and result in a power vacuum was Italy (King). The Weimar Republic soon followed, after Adolf Hitler gained people's trust and pushed others to the bottom (“Weimar Republic Definition, History, Constitution, Problems, Downfall, & Facts | Britannica.”). Shortly afterward, Czechoslovakia followed, then Denmark, Norway, Luxembourg, the Netherlands, Belgium, France, then finally the Philippines (Hart). In those instances of crisis, we saw that the world would either fall one way or the other, like a balance scale - there was limited stable middle ground.

So the question still stands: Do we currently face a democratic crisis? My answer to this is no; however, our society is slowly approaching a crisis through democratic backsliding based on patterns that resemble historic situations of democratic collapse and democratic erosion. We now stand close to the cliff -our choices that will determine the direction of future generations.

World War II

Returning to the example of WWII we see that it was a prominent example of democratic collapse that led to democratic crisis because our world had reached a tipping point. Before Hitler had risen to power, the Weimar Republic had multiple political parties such as the National Socialist German Workers Party (Nazis), but one thing that the Nazis had over other groups was the Sturmabteilung (SA) which loosely translates to stormtroopers - often known as brownshirts (Land). This group often beat up political enemies, people who didn't fit Nazi ideals, and anyone seen as a threat. In America, some critics argue that the growing authority and militarization of groups such as ICE reflect a broader expansion of executive power that has historically preceded democratic collapse. Today, ICE agents arrest and deport undocumented immigrants. In 2025, following the start of the second Trump Administration, reports of increased violence from ICE agents became apparent ("How ICE Went Rogue: Analysis of the Legal Authorities Governing Ice."). The most famous attack was in Minnesota on January 7th, 2026, when Renee Nicole Good was shot and killed after attempting to protect illegal immigrants from arrest with her car. After being asked to move several times, she started to turn her car but was shot anyway by an ICE agent that claimed she was trying to hit him (Musa). Critics have even claimed that ICE arrests are based on racial profiling. Historically, democratic erosion has accompanied the expansion of armed forces under the executive branch, such as ICE agents. However, the Nazi regime allowed the SA to utilize violence and fear from the beginning, killing and murdering anyone in their path. Whereas, ICE agents have historically been discouraged from violence and rarely used it up until the Trump administration took power. Recently there has been more pushback on ICE agents from the public due to media coverage and the checks and balances system. However, if left unchecked military powers such as ICE could continue to gain power and potentially utilize more violence against citizens. This is not a coincidence but instead a pattern of the expanding executive enforcement through military power that has historically preceded democratic crises.

Another key signal of democratic backsliding that could lead to a crisis is the executive branch bypassing congressional approval. In the months since Trump has taken office for his second term, he has bypassed gaining congressional approval for starting a war with Iran, invoking wartime powers to approve massive weapon sales to Middle Eastern allies without congressional oversight, stripped companies of funding previously provided to them by Congress, and issued multiple executive orders to allow him to change federal policies without needing legislative votes ("2025 Donald J. Trump Executive Orders."). These examples are worrisome because they offset the checks and balances system that was created to avoid autocracies. By strengthening the executive branch, it weakens legislative oversight, and dilutes the power of ordinary citizens who voted for their representatives in government. Our world is fast approaching a democratic precipice and could be pushed off the cliff into crisis.

Our Current Stance

At this point you may be asking yourself: How would we know if we are in crisis? It might help to look at an example of a similar point in time when we were at such a precipice: the Cold War. The Cold War, unlike WWII, represented a time of democratic erosion rather than collapse leading up to a crisis. Democratic erosion is defined as the lengthy process in which democratic norms, institutions, and processes deteriorate, usually due to internal sources such as the government (Williamson). Democratic erosion closely resembles modern day democratic backsliding. Though both words are often used interchangeably, they differ slightly depending on what forces are causing the deterioration of democratic norms. After World War II ended the world was rebuilding itself, but tensions between the USSR and the US were at an all-time high. Both countries sought to avoid major conflicts at all costs, leading to proxy wars and growing fear. However, these proxy wars were not as effective in avoiding major conflict as both countries would have hoped, placing pressure on democratic institutions. The Eastern European countries were the first ones that fell into communism's grasp during the Cold War, including Poland, Hungary, Czechoslovakia, Bulgaria, and Romania ("Fall of Communism in Eastern Europe, 1989."). Growing tensions in Korea led to a split where North Korea became a communist country, supported by the Soviet Union, and South Korea was a democracy, supported by the US. This conflict ultimately led to the bloody Korean War. The world unraveled slowly at first, then all too fast. We watched country after country fall, while attempting to prevent more democratic erosion. Yet it seemed we had fallen from the cliff and were spinning wildly out of control.

During the second half of the 20th century, fear was spreading throughout the world, not only because the two superpowers were at each other's throats, but also because of the fear of the domino theory. This theory suggests that if one country or nation in a region fell to communism, this could trigger other nearby nations to do the same ("Domino Theory: Definition, Cold War & Vietnam War.") Not only did democratic erosion occur through the collapse of past democratic nations but also through restrictions on civil liberties. Some examples include allowing the FBI to illegally bug private homes and offices, placing limitations on where citizens can travel, loyalty oaths became common in certain jobs, neighbors were afraid to share political stances, and the media seemed to continuously showcase anti-communistic propaganda.

This example of global politics during the Cold War era highlights that if democracy was truly in crisis, we would see countries falling, one after another. Our current world has not reached that point yet. However, just because democracy is not in crisis does not mean we remain unscathed, our democratic institutions still face large instances of democratic erosion and will continue to if left unchecked, based on past historic events.

On The Edge

Some may argue that our world is entirely free from crisis, that we stand nowhere near the path we had once taken over the last century. They have argued that major democracies have not fallen during large conflicts nor bloody wars since the Cold War. However, democracies do not normally collapse in sudden and bloody ways, instead they deteriorate gradually, and without

proper surveillance they will fall into autocracy. So yes, major democracies have not fallen during large conflicts yet they are still falling. For example, Hungary transitioned to an illiberal state without military intervention. Turkey fell into autocracy after an election. Russia remains an authoritarian regime with a democratic façade. Venezuela, similarly to Turkey, fell into autocracy after an election. ("Authoritarian Backlash: How Turkey and Venezuela Differ.") To others, little appears to change in America; however, democratic backsliding persists. We have become complacent with other countries' situations because we do not face their hardships, nor see their struggles. The idea of staying silent because it does not affect us is known as the Universe of Obligation, a phrase coined by sociologist Helen Fein ("Universe of Obligation | Facing History & Ourselves."). Though the Universe of Obligation was utilized by the Nazis we face a juxtaposition, instead of the government shielding our eyes from rapid democratic collapse, we are shielding our own eyes from the slow democratic backsliding that occurs around us. We walk a similar path as we once did, yet ours is different, democratic crises come in different forms, so it can become easy to believe that our situation is different. Yet history suggests that these smaller slower patterns can accelerate if we continue to leave them unchecked.

Others have argued that we stand nowhere near the path to crisis here in the US, that President Donald Trump is fully within his presidential rights to take the actions that he has. While this might be technically true, other countries have deemed America an "electoral democracy" losing the status of a "liberal democracy" for the first time in over 50 years (McKenzie and Carothers). An "electoral democracy" is a government that ensures citizens have the power to vote for government officials; however, a "liberal democracy" takes it a step further by ensuring citizens' personal freedoms. This loss of the title was attributed to concerns over the loss of civil liberties Americans have endured since the start of the second term of the Trump administration. Though this seems like an irrelevant detail, this follows the pattern of deterioration of democratic norms and principles through government agencies, and in this case through the executive branch. In fact it was found by Pew Research Center that in 2024 only 19% of Americans still believe the US is a good example of a democracy for other countries to follow (Janell). Furthermore in recent years the Economist Intelligence Unit has categorized the US as a "flawed democracy", falling just short of a "full democracy" ("Democracy Index 2022.").

Conclusion

These events, though individual, create an unsettling image of our future. If we do not attempt to prevent those with communist and authoritarian ideals from taking power, then our world will fall into democratic crisis. We face democratic backsliding, but as showcased by the Cold War and World War II, if we remain complacent, we allow today's small problems to accelerate.

We walk the path that we have walked a million times, and if we do not stop to turn around we will only fall into the trap. Generations have passed since our world has seen a

democratic crisis and the lessons have faded with time, however as George Santayana once said, “Those who cannot remember the past are condemned to repeat it.”

Works Cited

"Authoritarian Backlash: How Turkey and Venezuela Differ." Duke Today, 11 Sept. 2017, today.duke.edu/2017/09/authoritarian-backlash-how-turkey-and-venezuela-differ.

Bermeo, Nancy. "On Democratic Backsliding." Journal of Democracy, Jan. 2016, www.journalofdemocracy.org/articles/on-democratic-backsliding/.

Bradley W. Hart, PhD. "The Allies of World War II" Published August 29, 2024. https://www.nationalww2museum.org/war/articles/allies-world-war-ii.

Carrier, McKenzie, and Thomas Carothers. "U.S. Democratic Backsliding in Comparative Perspective." Carnegie Endowment for International Peace, 4 Dec. 2025, carnegieendowment.org/research/2025/08/us-democratic-backsliding-in-comparative-perspective.

Cartwright M. "Athenian democracy." World History Encyclopedia. Published online February 4, 2025. https://www.worldhistory.org/Athenian_Democracy/

"Crisis." Cambridge Dictionary, 1999, dictionary.cambridge.org/us/dictionary/english/crisis#google_vignette.

"Democracy Index 2022." Economist Intelligence Unit, 5 Jan. 2026, www.eiu.com/n/campaigns/democracy-index-2022/.

"Domino Theory: Definition, Cold War & Vietnam War." History.com, A&E Television Networks, 30 June 2025, www.history.com/articles/domino-theory.

"Fall of Communism in Eastern Europe, 1989." U.S. Department of State, U.S. Department of State, history.state.gov/milestones/1989-1992/fall-of-communism.

Fetterolf, Janell. "72% of Americans Say the US Used to Be a Good Example of Democracy, but Isn't Anymore." Pew Research Center, Pew Research Center, 10 July 2024, www.pewresearch.org/short-reads/2024/07/10/72-of-americans-say-the-us-used-to-be-a-good-example-of-democracy-but-isnt-anymore/.

"How ICE Went Rogue: Analysis of the Legal Authorities Governing ICE." American Immigration Council, 14 May 2026, www.americanimmigrationcouncil.org/fact-sheet/ice-cbp-legal-analysis/.

King, Russell S. "Italy - Fascism, Mussolini, Unification" Britannica, 23 May 2026, www.britannica.com/place/Italy/The-Fascist-era.

Land, Graham. "The Brownshirts: The Role of the Sturmabteilung (SA) in Nazi Germany." History Hit, History Hit, 28 Sept. 2021, www.historyhit.com/hitlers-bullyboys-the-role-of-the-sa-in-nazi-germany/.

Musa, Amanda. "Mother of 3 Who Loved to Sing and Write Poetry Shot and Killed by ICE in Minneapolis." CNN, Cable News Network, 10 Jan. 2026, www.cnn.com/2026/01/08/us/renee-nicole-good-minneapolis-ice-shooting-hnk.

"2025 Donald J. Trump Executive Orders." Federal Register, 2025, www.federalregister.gov/presidential-documents/executive-orders/donald-trump/2025.

"Universe of Obligation | Facing History & Ourselves." Facing History & Ourselves , 5 Jan. 2017, www.facinghistory.org/resource-library/universe-obligation.

Shapiro, Ian. “Democracy | Definition, History, Meaning, Types, Examples, & Facts” Britannica, 23 May 2026, www.britannica.com/topic/democracy.
“Weimar Republic | Definition, History, Constitution, Problems, Downfall, & Facts” Britannica, 22 May 2026, www.britannica.com/place/Weimar-Republic.
Williamson, Vanessa. “Four Things to Know about Democratic Erosion.” Brookings, 18 Oct. 2023, www.brookings.edu/articles/four-things-to-know-about-democratic-erosion/.

Simulating Stability: How Period Ratio Complexity Affects Orbital Resonance Lock in a Computational N-Body Model By Aarib Mahd Syed

Abstract

Orbital resonance is when two or more planetary objects exert gravitational forces on each other, resulting in their orbital periods forming a simple ratio, like 2:1 or 3:2. It matters because these interactions stabilize planetary systems, prevent collisions, and shape the structure of systems from moons to exoplanets.

To investigate orbital resonance, I developed a computational physics engine using the AI-assisted coding tool Cursor. The simulation modeled multiple resonance configurations and measured lock stability under different conditions.

I conducted 60 simulation runs testing 2:1, 3:2, and 4:3 resonance configurations (15 runs each), plus 15 near-resonance drift tests.

Results showed that in this simulation, configurations with more complex period ratios (4:3) achieved lower lock confidence scores than simpler ratios (2:1), with confidence declining from 98% to 53% as ratio complexity increased. These findings reflect how the simulation's detection metric responds to different period ratio configurations, and they provide a basis for further computational study of resonance behavior in planetary systems.

Introduction

Orbital resonance is a gravitational phenomenon in which two or more orbiting bodies repeatedly align their orbital periods, forming ratios such as 2:1 or 3:2. These aligned meetings allow the bodies to periodically use gravitational forces on each other, which rhythmically reinforces their motions (“Orbital Resonance”).

Over longer periods of time, resonance can stabilize orbits and prevent close encounters or collisions, which maintains the structure of systems from moons to packed exoplanet chains. These resonance patterns can exist for billions of years (“Orbital Resonance”).

This study examines how orbital bodies achieve and maintain resonance lock across different period ratios, comparing stability characteristics of 2:1, 3:2, and 4:3 configurations in a simulated environment.

After establishing the theoretical background, this paper describes the simulation’s development and validation, presents results from 60 test runs, and analyzes how period ratio complexity affects lock stability.

Background

As bodies in orbital resonance with each other meet at nearly the same orbital phases, each encounter produces a similar, periodic gravitational kick in angular momentum. One main sign of planetary resonance is phase locking, when the bodies’ gravitational interactions synchronize. Orbital resonance locks the timing of conjunctions, which prevents repeated close approaches at the wrong phases and maintains orbital alignment (Malhotra et al.). In certain

examples of resonance such as Neptune and Pluto, their 3:2 resonance prevents them from colliding on their overlapping orbital paths for billions of years (Malhotra; Williams and Benson).

The resonance ratio is a term that refers to the relationship that exists between the orbital periods of the two bodies in terms of integer fractions. When the resonance ratio is 2:1, this implies that the inner body makes precisely twice as many orbits as the outer body does in the same period of time. On the other hand, the orbital period refers to the time taken by a body to make a single orbit around its primary body, starting at a point of conjunction and reaching the point again on the other side of its primary body. Locking refers to the process in which the bodies maintain a resonance ratio while at the same time keeping their period ratio as close as possible to the desired integer fraction.

A good example of orbital resonance is the case of Jupiter's Galilean satellites. Io, Europa, and Ganymede are in a 1:2:4 orbital resonance, which is known as the Laplace resonance, with Ganymede's orbital period of 7.155 days being exactly twice that of Europa, which has an orbital period of 3.551 days, which is, in turn, exactly twice that of Io, which has an orbital period of 1.769 days. This means that, in the Jupiter system, each time Ganymede goes around Jupiter, Europa goes around Jupiter twice, and Io goes around Jupiter four times. This resonance causes significant tidal heating, with Io showing strong tidal activity, making it the most volcanically active body in the solar system. This gravitational synchronization has kept the orbit of Jupiter's moons stable for billions of years (NASA, "Ganymede in Depth"; "Orbital Resonances").

Kepler's Third Law provides the foundation of the mathematical model for orbital resonance. It states that the square of the orbital period T is directly proportional to the cube of the semi-major axis a, or $T^2 \propto a^3$. This law means that when bodies have specific ratios of orbital distances, their ratios of orbital periods remain constant. For example, in a 2:1 orbital resonance, the outer body will be about 1.587 times farther from the central body than the inner body. This is calculated as $2^{(2/3)} \approx 1.587$. In a 3:2 orbital resonance, the bodies will have a distance ratio of approximately 1.31, calculated as $1.5^{(2/3)} \approx 1.31$ (NASA, "Orbits and Kepler's Laws"; Murray and Dermott).

Kepler's First Law, which states that the orbits are not perfect circles but are instead elliptical, plays a part in resonance dynamics as well. Research has shown that the elliptical nature plays a vital part in the operation of the resonance restoring mechanism. The gravitational pull between objects meeting at different points on their elliptical orbit causes the transfer of angular momentum, which is not the same before and after the conjunction. This provides a natural balancing force that pushes the meeting point towards the equilibrium position along the axis of symmetry (Jiang; Murray and Dermott).

Kepler's Second Law, which states that equal areas are swept in equal times, is the reason why objects move faster in their orbits near periapsis and slower near apoapsis. The change in the rate of the objects has a bearing on the stability that will be achieved, depending on the resonance that is formed (NASA, "Orbits and Kepler's Laws").

Methods

I built an N-body gravitational simulation engine that models how celestial bodies orbit and interact with each other through gravity. This engine includes a resonance analyzer that detects whether or not the orbiting bodies fall into a specific resonance. The system uses a Python backend that handles all the physics calculations and a React frontend that provides a proper visualization to watch each planet's orbit, add or remove bodies and see the resonance data update as the simulation continues in real time.

The backend is written in Python using the Fast API framework, which provides a REST API. and the server runs on unicorn and exposes endpoints for controlling the simulation, loading presets, and exporting data. The frontend is built with React and JavaScript, which provides a visualization for the orbiting bodies. Communication between the frontend and backend happens through the REST API for streaming the real-time positions and data.

The simulation uses Velocity Verlet to solve the equations of motion, through a symplectic integrator, which is much more efficient in energy conservation, especially over longer periods of time when compared to simpler methods like Euler integration (Quantum Tinkerer).

The simulation does multiple calculations to accurately collect and display data for each planet. These include Newton's Law of Universal Gravitation, orbital period detection (θ = atan2(dy, dx)), period ratio and resonance detection (Period Ratio = T_outer / T_inner, Deviation = |Actual Ratio - Target Ratio| / Target Ratio × 100%, and Drift Rate = σ_period / T_average), total mechanical energy (Kinetic Energy = ½ × m × v², Potential Energy = -G × M × m / r, and E_total = $\Sigma(\frac{1}{2}m_i v_i^2) - \Sigma(G \times m_i \times m\square / r_i\square)$), circular orbit velocity ($v = \sqrt{(G \times M / r)}$), and tidal forces ($a_{tidal} \propto (M \times R^5) / (Q \times r^8) \times v_{relative}$).

I used multiple ways of validating the simulation. First, the code computes the total mechanical energy of the system (kinetic energy plus gravitational potential energy) at every timestep. In a properly working gravitational simulation, this total energy should remain approximately constant over time, which it does. I also compared the resonance patterns and deviation rates to real world examples.

I have carried out experiments with four preset configurations. The 2:1 Resonant Pair configuration consists of two planets at distances from the Sun determined by Kepler's Third Law. The result is that the orbital period of the outer planet is twice that of the inner planet. The 3:2 Resonant Pair configuration is achieved in a similar manner with a period ratio of 1.5, similar to the Neptune-Pluto resonance. The 4:3 Resonant Pair configuration is achieved with a slightly higher resonance rate, where the outer body makes three orbits for every four made by the inner body. The Resonance Drift configuration simply has the outer body placed at 8% distance from the 2:1 position.

A number of key metrics were collected in each simulation. Confidence is a percentage from 0% to 100%, indicating the stability and reliability of the resonance lock. A higher percentage means a tighter lock. Deviation is specified as a percentage indicating how much the actual measured period ratio varies from the closest integer resonance target. Drift rate is

specified as the rate of change of the period ratio over time. It is calculated as the standard deviation of the measured orbital periods divided by the mean period. Status is a categorical label indicating whether the measurement is LOCKED (stable resonance achieved), TRANSITIONAL (moving towards stable resonance), NOT LOCKED (no resonance achieved), or MEASURING (not enough orbits completed to determine status). The results in this test all fell into either LOCKED or NOT LOCKED.

The resonance analyzer checks to see if the bodies are in resonance using the following three criteria, which must be satisfied simultaneously. First, the actual ratio must be within 2% of the nearest integer ratio. For example, the ratio of 2.015 has a deviation of 0.75% from the integer ratio of 2:1. Second, the drift rate must be less than 3%, which means the orbital periods are stable and do not change significantly with time. Finally, both objects must have completed at least three full orbits. If all three conditions are satisfied, the status will be LOCKED. The confidence level will then be computed as 1 minus the ratio of the actual deviation to the 2% threshold. For example, an actual deviation of 0.08% will yield a confidence level of 96%. If the actual deviation is between 2% and 5%, with a corresponding drift rate of less than 5%, the status will be TRANSITIONAL. Otherwise, the status will be NOT LOCKED with 0% confidence.

I conducted a total of 60 simulation runs for the main experiment. This consisted of running each of the four presets (2:1, 3:2, 4:3, and Resonance Drift) exactly 15 times each to gather statistical data. Additionally, I ran 15 separate mass effect tests on the 2:1 preset, changing the planet masses to see how mass ratios affect resonance stability.

Each simulation ran until bodies completed at least 5 full orbits after achieving lock status, or for a minimum of 50 orbits if lock was not achieved.

Because the simulation is deterministic and uses fixed initial conditions, each of the 15 runs per configuration produces the same result. The 15 runs were used to check that the simulation was consistent, not to generate independent data points, which means the effective sample size is one result per configuration rather than fifteen.

Results

In this simulation, I observed clear and consistent patterns across 60 total runs testing four different resonance configurations. Each of the four presets (2:1, 3:2, 4:3, and near-resonance drift) was run 15 times under identical conditions, with each simulation allowed to run for a maximum of 150 time units or until the system achieved resonance lock and completed at least 20 orbits. The results demonstrated a direct relationship between the complexity of the period ratio and the strength of the resulting resonance lock.

2:1: The simplest type of resonance, where the inner body orbits twice for every one orbit of the outer body, has the tightest lock. The deviation from the ideal ratio here is only 0.077%, and the confidence factor is 98.1%.

3:2: The second type of resonance, where the inner body orbits three times for every two orbits of the outer body, which keeps Neptune and Pluto from colliding, has a weaker lock than

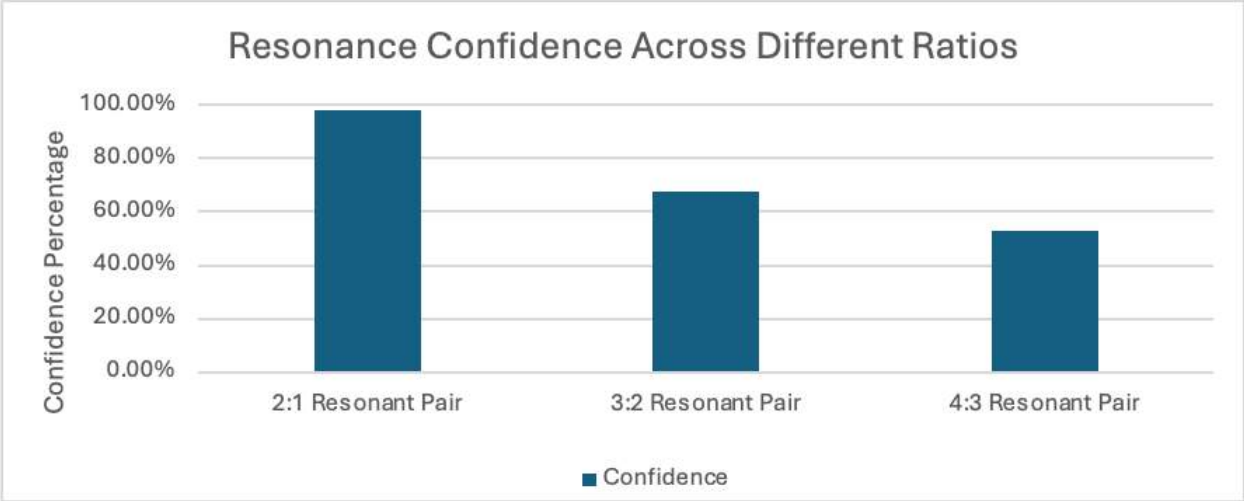
the 2:1 configuration. The deviation from the ideal ratio here is 0.973%, and the confidence factor is 67.8%.

4:3: The most complex type of resonance, with the inner body orbiting four times for every three orbits of the outer body, has the weakest lock among the three. The deviation from the ideal ratio here is 1.265%, and the confidence factor is 53.0%.

Table 1: Main Simulation Results Across Resonance Configurations

Preset	Target Ratio	Actual Ratio	Deviation %	Status	Confidence	Drift Rate
2:1 Resonant Pair	2:1	2.0008	0.08%	LOCKED	98.10%	0.000198
3:2 Resonant Pair	3:2	1.5097	0.97%	LOCKED	67.80%	0.001453
4:3 Resonant Pair	4:3	1.346	1.27%	LOCKED	53.00%	0.001487

Figure 1: Confidence Results Across Resonance Configurations

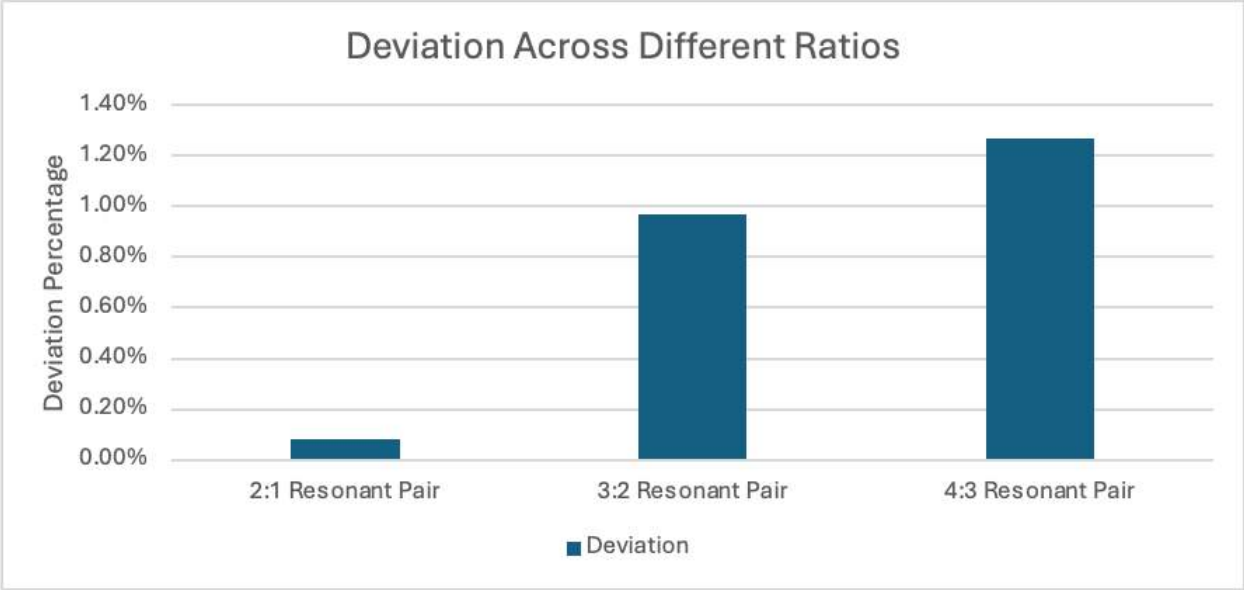


In contrast, the near resonance drift, which started at 8% deviation from the ideal 2:1 ratio, failed to achieve resonance at all during the 15 runs. The deviation from the ideal ratio remained constant at 24.37%, and the confidence factor remained at 0% at all times.

Table 2: Results from Near-Resonance Drift Tests

Preset	Target Ratio	Actual Ratio	Deviation %	Status	Confidence	Drift Rate
Resonance Drift	Near 2:1	2.2437	24.37%	NOT LOCKED	0.00%	0.000121

Figure 2: Deviation Results Across Resonance Configurations



In this simulation, it was observed that decreasing the mass of the outer planet produced the best possible results, where the outer-mass 0.5x simulation had a perfect 0.00% deviation and 100% confidence. However, increasing the mass of the outer planet had a detrimental effect on resonance, where increasing it 10 times had only 19.7% confidence and a 3.27% deviation. For the inner planet, it was observed that increasing its mass 10 times had a less significant effect, where it had 90.1% confidence. However, the simulation where the inner planet was decreased 10 times and the outer planet was increased 10 times had the second-worst possible resonance at 22.9% confidence. This observation is consistent with actual planetary bodies where resonant moons are smaller compared to their parent planets (NASA, “Ganymede in Depth”; D. Williams). This is likely because a more massive outer body exerts stronger gravitational perturbations on the inner body, which increases how much the measured period ratio deviates from the ideal value. The fact that reducing the outer planet's mass to 0.5x produced a 100% confidence score suggests that at low enough mass ratios, the perturbations from the outer body become small enough that they no longer affect the lock quality.

Table 3: Mass Variation Test Results on the 2:1 Preset

Run #	Test Name	Inner Mass	Outer Mass	Actual Ratio	Deviation	Status	Confidence	Drift Rate
1	Baseline	1.0x	1.0x	2.0008	0.08%	LOCKED	98.10%	0.000198
2	Inner 0.5x	0.5x	1.0x	2.0008	0.08%	LOCKED	98.00%	0.000142
3	Inner 2x	2.0x	1.0x	2.0007	0.07%	LOCKED	98.10%	0.000419
4	Inner 5x	5.0x	1.0x	2.0011	0.11%	LOCKED	97.20%	0.001243

5	Inner 10x	10.0x	1.0x	2.004	0.40%	LOCKED	90.10%	0.003564
6	Outer 0.5x	1.0x	0.5x	2	0.00%	LOCKED	100.00%	0.000171
7	Outer 2x	1.0x	2.0x	2.0028	0.28%	LOCKED	92.90%	0.000386
8	Outer 5x	1.0x	5.0x	2.0126	1.26%	LOCKED	68.80%	0.00118
9	Outer 10x	1.0x	10.0x	2.0327	3.27%	LOCKED	19.70%	0.00198
10	Both 2x	2.0x	2.0x	2.0032	0.32%	LOCKED	92.00%	0.000736
11	Both 5x	5.0x	5.0x	2.0177	1.77%	LOCKED	56.00%	0.003607
12	Inner 0.1x	0.1x	1.0x	2.0009	0.09%	LOCKED	97.80%	0.000134
13	Outer 0.1x	1.0x	0.1x	1.9995	0.05%	LOCKED	98.80%	0.000147
14	Inner 10x, Outer 0.1x	10.0x	0.1x	1.9994	0.06%	LOCKED	98.40%	0.002289
15	Inner 0.1x, Outer 10x	0.1x	10.0x	2.0313	3.13%	LOCKED	22.90%	0.001909

Discussion

From the results, it is evident that within this simulation, and as measured by the lock confidence metric, more complex resonance ratios produce lower confidence scores, with the 2:1 configuration reaching 98.1% and the 4:3 configuration reaching only 53.0%. The 2:1 configuration, which represents the simplest form of resonance for two orbiting bodies, achieved 98.1% confidence with a deviation of 0.077% from the ideal ratio. The 3:2 configuration achieved 67.8% confidence with a deviation of 0.973%, while the 4:3 configuration achieved 53.0% confidence with a deviation of 1.265%. The near resonance drift test, which was set up at the wrong orbital distance, failed to achieve a lock.

There are various theories that explain the decrease in confidence as the complexity of the ratio increases. First, the sensitivity of the measurement would be a major contributing factor. The 2:1 ratio has a larger difference from the next closest ratio, which makes it more difficult for the measured ratio to fall beyond the 2% threshold required for the analyzer to lock into the orbit. The 4:3 ratio, which has a measured ratio of 1.333, has adjacent ratios such as 5:4 with a measured ratio of 1.25 and 3:2 with a measured ratio of 1.5.

Secondly, the amount of time used for the stabilization process may have played a role in the results. The simulations have a maximum of 150 time units. Although all three locked configurations achieved lock in 19 to 27 time units, the 4:3 ratio may need to have a longer amount of time to achieve lock because this will give the ratio a chance to get closer to the ideal ratio. In actual celestial systems, the resonance has been present for billions of years to achieve equilibrium.

Thirdly, the initial conditions used in the simulation may have played a role in the results. Although the preset positions place the planets at a distance from each other that is based on Kepler's Third Law to achieve the desired ratio, the actual interactions in the simulation use simplified two-body interactions with a central star. The gravitational interactions between the two orbiting planets have slight perturbations in their orbits. These perturbations may have a stronger effect on the 4:3 ratio because the planets will be closer to the star. Fourthly, because both planets are initialized at orbital distances derived from Kepler's Third Law for each target ratio, the resonance configuration is already built into the starting conditions. This means the simulation measures how well a nearly perfect resonance is maintained over time, rather than whether resonance forms naturally, which is how resonance actually develops in real planetary systems.

However, it must be noted that the real-life 3:2 and 4:3 resonances show remarkable stability over time. The 3:2 resonance between Neptune and Pluto has lasted over four billion years, with barely any deviation (Malhotra; Williams and Benson). The Galilean moons of Jupiter have kept their 1:2:4 Laplace resonance since the beginning of the solar system's development (NASA, "Ganymede in Depth"; "Orbital Resonances"). The Titan-Hyperion system in the Saturnian ring system has kept its 4:3 resonance with a deviation of only 0.08% (Cornell University). These real-life examples show that the 3:2 and 4:3 resonances are not less stable than 2:1 resonances in actual planetary systems, which means the declining confidence values in this simulation are more likely a product of the detection methodology than a reflection of an underlying physical difference. This is likely because the detection algorithm evaluates stability over 150 time units, which may not be enough time for more complex ratios to settle near their ideal values, given that real systems have had billions of years of tidal evolution to reach the deviations that are currently observed.

In this simulation, the deviation is 0.077% for the 2:1 configuration. This can be compared with real-life astronomical systems, such as the Enceladus-Dione system in the Saturnian system, which is in 2:1 resonance with an observed deviation of 0.12% (Cornell University). In the 3:2 configuration, the deviation is 0.973%, which is higher than the observed 0.29% in the Neptune-Pluto system (Willman; D. Williams). This is likely because the Neptune-Pluto system has had time to get into closer resonance during its very long evolutionary time. Finally, in the 4:3 configuration, the deviation is 1.265%, which is much higher than the observed 0.08% in the Titan-Hyperion system (Cornell University). Overall, these comparisons suggest that the simulation can mimic the general behavior of orbital resonance, but it overestimates the deviations in systems where the resonance is more complex. It is likely that this is because the model does not capture the gradual tidal evolution these systems have had over billions of years.

If I were to continue this research, I would have several other tests that could be carried out. First, I would greatly extend the simulation period past 150 time units and see if it is possible for these 3:2 and 4:3 resonances to lock more closely over a longer period of time, potentially showing whether these differences are due to a limitation of the simulation process or

a true difference between these resonances. Second, it would be possible to simulate other resonances, for example, 5:3, 5:4, and 7:4, and see if there is a continued linear progression of decreasing confidence scores. Thirdly, I would add in more data and possibly conduct more research on different systems during the making of the simulation to possibly produce a more accurate physics engine. Lastly, it would be possible to add randomized initial conditions, for example, small variations in starting positions and speeds, and see if it is possible for the simulation to produce different results each time it is run rather than identical results, potentially providing a more realistic set of data with meaningful standard deviations.

Conclusion

In conclusion, in this project I aimed to determine how orbital bodies achieve resonance lock by creating and testing a computational model of an N-body simulation. Out of 60 resonance tests (2:1, 3:2, 4:3, and near-resonance drift), and 15 additional tests on varying masses, the model indicated that, within this simulation and as measured by the lock confidence metric, simpler period ratios produced stronger lock scores than more complex ones.

In general, as the complexity of the resonance ratio increased, the lock confidence score decreased, with confidence falling from 98.1% at 2:1 to 67.8% at 3:2 and 53.0% at 4:3. The near-resonance drift configuration failed to achieve lock at all, which shows that resonance is not guaranteed when bodies are placed at the wrong orbital distances. The mass variation tests further indicated that higher outer-planet mass reduces lock confidence, which is consistent with real planetary systems where resonant moons are significantly less massive than their host planets.

At the same time, the study does raise some questions about how much of this trend is physically driven, versus how much is a product of the simulation environment's underlying assumptions. For example, the real resonance systems such as Neptune-Pluto (3:2) and Titan-Hyperion (4:3) seem to be held with greater stability than the simulation predicted, which may be a product of the relatively short time period that the simulation ran (150 time units), which may not be sufficient to allow higher-order ratios to fully settle into their patterns. Furthermore, the fact that the simulation environment is deterministic, with fixed initial values, means that running the simulation again with the exact same settings will always produce the exact same result, which limits the statistical power of the study.

Future work could seek to fix these issues by incorporating randomized initial values to allow a true statistical study, running the simulation over a longer time period to see if the 3:2 and 4:3 ratios trend towards higher confidence over time, and expanding the study to more complex ratios such as 5:4 or 7:4 to see if the trend towards instability is linear in nature. It could also be useful to extend the study to multi-resonant chains, such as the TRAPPIST-1 system (NASA, "Astronomers Confirm Orbital Details of TRAPPIST-1h"), to see how the stability of resonance might change in crowded environments. Finally, comparing the simulation environment's output directly to real-world mission-based datasets such as the Juno and Cassini datasets could help to tie the study more directly to the real world.

Works Cited

Cornell University. “Rings Workshop.” Cornell University Department of Astronomy, 2012, https://hosting.astro.cornell.edu/specialprograms/reu2012/workshops/rings/.

Jiang. “Physics in Orbital Resonance.” Stanford University, 2007, http://large.stanford.edu/courses/2007/ph210/jiang2/.

Malhotra, R. “The Origin of Pluto’s Orbit: Implications for the Solar System Beyond Neptune.” The Astronomical Journal, vol. 110, 1995, pp. 420-429, https://ui.adsabs.harvard.edu/abs/1995AJ....110..420M/abstract.

Malhotra, R., et al. “Chaos and Stability of the Solar System.” Proceedings of the National Academy of Sciences, 2001, https://www.pnas.org/doi/10.1073/pnas.231384098.

Murray, C. D., and S. F. Dermott. Solar System Dynamics. Cambridge University Press, 1999, https://assets.cambridge.org/97805215/72958/sample/9780521572958wsn01.pdf.

NASA. “Astronomers Confirm Orbital Details of TRAPPIST-1h.” NASA Science, https://science.nasa.gov/universe/exoplanets/astronomers-confirm-orbital-details-of-trappist-1h/.

NASA. “Ganymede in Depth.” NASA Solar System Exploration, https://solarsystem.nasa.gov/moons/jupiter-moons/ganymede/in-depth/.

NASA. “Orbits and Kepler’s Laws.” NASA Science, https://science.nasa.gov/solar-system/orbits-and-keplers-laws/.

“Orbital Resonance: The Dance of Heavenly Bodies in Synchrony.” EarthSky, https://earthsky.org/space/orbital-resonance-dance-heavenly-bodies-synchrony/.

“Orbital Resonances of Galilean Moons.” The Planetary Society, https://www.planetary.org/space-images/orbital-resonances-of-galilean-moons.

Quantum Tinkerer. “Molecular Dynamics: Week 3.” Delft University of Technology, https://compphys.quantumtinkerer.tudelft.nl/proj1-moldyn-week3/.

Williams, D. R. “Planetary Fact Sheet.” NASA Space Science Data Coordinated Archive, https://nssdc.gsfc.nasa.gov/planetary/factsheet/.

Williams, J. G., and G. S. Benson. “Resonances in the Neptune-Pluto System.” The Astronomical Journal, vol. 76, 1971, p. 167, https://articles.adsabs.harvard.edu/pdf/1971AJ.....76..167W.

Willman, B. “Planetary Systems: Sol.” Princeton University, https://www.princeton.edu/~willman/planetary_systems/Sol/.

Play Vs. Academics in Early Childhood By Isabel Hershkin

Fred Donaldson, a Ph.D. play specialist, once stated that, "Children learn as they play. More importantly, in play, children learn how to learn." The debate between play-based learning and academic training has shaped early childhood education for decades. Generally, academic training emphasizes memorization, practice, and teacher-led instructions aimed at mastering different school subjects. However, developmental theorists argue that play is the 'work' of childhood. Many of these theorists argue that children are not passive learners; rather, they are active builders of knowledge who learn best through exploration while using gross and fine motor skills. Moreover, many credible organizations also now emphasize that play is vital for healthy brain development and long-term academic success. Play is the most essential concept to early childhood, as compared to academic training, because it acts as the main notion for learning through emotional regulation, physical health, and social skills.

Unlike academic schooling, self-directed play acts as the primary platform for children to develop strong emotional regulation. Emotional balancing is one of the most critical concepts embraced during early childhood development. This is because it enables children to manage their intense emotions starting from an early age, improving their academic readiness, and behavioral control. Emotional growth in play-based learning also allows children to experience a variety of situations. It also helps them manage frustration, develop empathy, and maintain focus as they play either independently or as a group. For example, the importance of emotional regulation in child-based play states, "A wealth of research [that] shows that unstructured play—play that isn't organized or directed by adults or older peers and that generally doesn't have a defined purpose or outcome—is a fundamental necessity for children to thrive . . . emotionally [and] mentally" (Clay, 2023). This suggests how play-based learning during early years relies on emotional regulation to enable children to shift from being overwhelmed by emotions to exploring and learning. It also reveals how emotional control contributes to academic learning through children concentrating on tasks for a longer period of time and by boosting their learning abilities and outcomes. Furthermore, integrating play into daily life can foster higher emotional regulation in young children. According to Monmouth University psychology professor Robyn Holmes, Ph.D., "By incorporating playfulness into everyday tasks like cooking dinner or planting a garden, parents will not only be getting help, they will be teaching children important skills like math and collaboration, and giving them a sense of contributing to the household" (Clay, 2023). This emphasizes the idea that integrating play-based learning into daily routines allows children to develop crucial skills while fostering a sense of contribution. This approach can also reduce anxiety and promote emotional growth by transforming chores into engaging learning opportunities. The positivity brought by emotional regulation in early childhood development is one of the core principles in the argument for play-based learning over academic training in childhood.

Active, unstructured play in early childhood is superior to academic drills in promoting physical well-being through direct engagement with a child's environment. Physical health is another key notion in the argument for play over academics. This is because it strongly develops

both fine and gross motor skills by prioritizing play over intense academics. Therefore, incorporating play and movement is necessary for children to build healthy minds and bodies, which ultimately supports their long-term learning and well-being. This evidence further demonstrates how play, especially physical play, is being taken away from children starting at a young age. According to the National Association of Elementary School Principals, "96% of surveyed school systems had at least 1 recess period. Another survey a decade later found that only 70% of even kindergarten classrooms had a recess period" (Ginsburg, 2007). This data is important because it shows that recess time kindergarten has decreased and how it has dropped by 26% in a decade. This is extremely significant because being physically active is valuable to children, especially as they begin to build their energy and agility at a younger age. Moreover, physical well-being in childhood is most effective when used in a child-centered environment. This evidence has been specifically proven, stating that, "Even a formal structured physical education class may not offer the same benefit as free-play recess" (Ginsburg, 2007). These findings highlight the idea that physical education in school is not as beneficial for growing children as free, child-initiated physical activities. Consequently, children often feel freer and more engaged when doing unstructured activities, as they are in control of their movement and play choices. The liveliness that physical health brings to children at an early age is an extremely important concept in the debate over whether play-based learning is superior to growing children.

Play-based learning acts as a vital, natural principle for early childhood social development, allowing children to develop their social skills through peer engagement in play. Social skills are another significant concept included in play-based learning at an early age. This is because it allows children to develop cooperation and empathy in a safe space. Through play, children also learn to navigate relationships and manage their emotions, which ultimately builds critical social competence. At Tel Aviv University, researcher and professor Sara Smilansky states how, ". . . children who show the greatest capacities for social make-believe play also display more imagination and less aggression, and a greater ability to use language for speaking and understanding others" (Almon, 2003). This highlights how significant social play is at a young age and how it can actually lessen aggression in children and expand their imagination further. This also explains how social play supports social skill development, with children interacting with each other more and having a greater ability to use language for speaking and understanding their peers. Even without being able to understand each other, children are brought together under the common domain of play. Research clarifies how, ". . . fascinating [it is] to watch children from different countries playing together. Although they may not be able to speak one word of the other's language, they can play together for hours. They enter a common realm where the external differences of language and culture are small compared to the vast similarities embedded in the child's inner urge to play" (Almon, 2003). This demonstrates how no matter what country they are from or what language they speak, children are able to find entertainment in play together. They are also able to create an environment where they can play freely together

and socialize with each other. The significance of social skills in play-based learning is another dominant factor in the dispute that includes play and academic achievements in early childhood.

Although play has been proven to be more effective than academics in children's long-term achievements, many still believe that academics are superior to play for growing children. Some benefits of academic training include improved language, literacy, math skills, and problem-solving, along with better self-regulation and increased motivation to learn. However, the play-based learning approach enhances these academic outcomes, such as math and literacy. This idea is supported by the claim that, ". . . advocates of play arguing for the return of play in early childhood education classrooms still have difficulty convincing the proponents of more rigorous academic instruction. These proponents cite research pointing to insufficient evidence of the positive effect of play on children's short- and longer-term social, emotional, cognitive, and behavioral outcomes as well as their overall well-being" (Bodrova et al.). While academic training is often viewed as more important than play-based learning, play has been proven to be essential to childhood development and is considered the 'work' of childhood. Although some critics argue that there is limited evidence supporting play-based learning, real-world experiences and research demonstrate that play offers greater short- and long-term benefits for children's overall well-being. Play is the most essential concept to early childhood, as compared to academic training, because it acts as the main notion for learning through emotional regulation, physical health, and social skills.

My topic of play-based learning being superior to academic training in early childhood development has taught me a great deal about the significance of play for children. While researching, my opinion on this topic has greatly deepened as my understanding of the benefits of play has grown. I have learned that these benefits, including social skills, physical development, and emotional regulation, are fundamental to overall growth in children, often fostering these skills more effectively than traditional academics in early childhood. The most challenging part of writing this paper was finding a balance between advocating for play-based learning and addressing the pressure for academic success, as I wanted to overcome the belief that play is just 'having fun' without learning goals. This debate is also important for people to understand because it highlights how play improves long-term academic, social, and emotional outcomes rather than just short-term school readiness. As Donaldson said, "In play children learn how to learn," reinforcing the idea that play is essential to meaningful and lasting early childhood development.

Is It Ever Wrong to Do the Right Thing for the Wrong Reason?: Morality Through the Kantian Perspective By Olivia Peng

Abstract

The question, *Is it ever wrong to do the right thing for the wrong reason?* is a long-debated ethical dilemma. To answer this question, we must turn to two philosophical frameworks used to determine whether an action would be right or wrong: Deontology and consequentialism, opposing beliefs that come together to answer the same moral question. In this paper, I argue that it is wrong to do the right thing for the wrong reason; based on Kantian deontology, an action cannot be morally right unless the intent is to act out of duty to universal moral law. If only the outcome was taken into account, people may act only out of self benefit which would deconstruct social relationships and weaken society.

Introduction

"It is the motive alone which gives real moral worth to the actions of men," wrote Jean de La Bruyère, who was a seventeenth-century French philosopher, known for *Les Caractères* (1688), a collection of observations and analyses that pointed out flaws of people in French high society and court. [1] Although many people only focus on whether an action proves beneficial, intentions reveal an individual's true character. Throughout the yellowing pages of history, numerous philosophers and authors have debated whether courtesies were still honorable in the case that they are done out of selfishness, greed, or the desire for attention. Two primary philosophical frameworks used to determine whether an action would be morally right or wrong are deontology and consequentialism; two sides of the same coin of justice, yet different in how they approach moral situations. Deontologists believe that actions are just if they align with universal rules of duty, while conversely, consequentialists believe that the morality of an action is determined solely by the outcome and that the intent is inconsequential. [2] A "right thing" is merely subjective; a deontologist would define it according to strict moral rules and duties, while in contrast, a consequentialist bases the "correct thing" by the results of the action, believing the best outcome justifies the means. [3] It can be wrong to do the right thing for the wrong reasons; though the immediate effect may be seen as positive, people may decide their path of action through harmful motives that can erode trust and that support negative values that may start to erode the foundations of society.

The Categorical Imperative

The eighteenth-century philosopher Immanuel Kant, known as "The Königsberg Clock", and famously one of the foremost thinkers of the Enlightenment, argued that good deeds lose their value when performed for selfish reasons. Kant developed an ethical framework, Kantian deontology, which argues that actions only have moral worth when done out of duty and respect for the law, away from the path that is motivated by self-interest. [4] Central to Kantian deontology is the idea that the morality of an action is determined by the principle behind it

rather than the produced consequence. Here is where the "categorical imperative" becomes important- it is a universal law that is "unconditional or absolute for all agents the validity or claim of which does not depend on any desire or end", and requires individuals to do what is right because it is what is expected of them, not to obtain a reward or personal benefit. [5] It serves as a supreme test of morality and rationality for an action. Under the guidance of this system, a good deed performed for selfish reasons would be lacking in genuine moral worth. [6] But why must intention matter? Why are morals important to everyday life? Morality is a key factor of humankind; defining not only the decisions of individuals, but also the foundational code of conduct for society to thrive. [7] People rely on moral frameworks to ensure cooperation, stability, and mutual understanding in communities. Without a shared sense of ethical intention, social relationships would simply be a transaction; a mutual exchange of value, that drains the authenticity of interpersonal connections. Kant's point of view suggests that when actions are driven by self-interest alone, they lose their integrity and risk undermining trust because it causes others to question whether their behavior is genuine, and fails to meet the standards of genuine moral behavior.

Critique of Corruption

Kant's philosophies may be applied towards numerous real-life situations. Consider a wealthy individual who donates large sums of money to charities, claiming it was out of their own obligation and sense of duty, when rather it was to gain social approval and spread the word of their "appealing" public image. Though the funding may help the organization, the individual's intent is selfish rather than compassionate. A contemporary example would be businessman Elon Musk's philanthropic activities through the Musk Foundation. On March 10, 2024, the New York Times first launched a major investigative report detailing how Elon Musk implemented his foundation to funnel billions of dollars to support causes directly beneficial to himself and only organizations in his personal orbit. He was scrutinized for neglecting his duty of philanthropic giving. [8] The IRS, the Internal Revenue Service, mandates that private foundations must distribute at least 5% of their non-charitable-use assets for charitable purposes to maintain a status of tax exemption. [9] It is noted that Musk's foundation has missed the deadline repeatedly. [10] The donations have had certain positive outcomes, such as expanding educational opportunities and supporting scientific initiatives, but the moral assessment changes if such acts are performed primarily for personal advantage. In these circumstances, the motivation matters significantly because it suggests charitable gifting is being used as a tool for self-interest rather than moral duty. Kant would argue that such actions would not possess much moral worth due to its main motivation of personal benefit over duty or collective good. [11] To illustrate this principle, Kant proposed the example of a shopkeeper who charges customers fair prices only to uphold their professional standing and generate commerce. [12] Although the shopkeeper acts honestly, Kant maintained that morality is not achieved due to the shopkeeper's incentive of personal gain and influence rather than a commitment to doing what is right. Likewise, if charitable giving is motivated by personal factors, then the action may be beneficial

to that individual and perhaps society, but cannot be considered genuinely moral by the standards of Kantian ethics.

Rockefeller and the Consequentialist Perspective

Some people would choose the path favoring outcome over intention; in the case an ill-intentioned action brings forth welcome results, then it shall be considered morally good. This perspective can be observed from the life of John D. Rockefeller, the founder of the Standard Oil Company and the first billionaire in modern history. [13] Rockefeller sought to acquire affluence, amass wealth, and dominate the oil industry. His goal was to eliminate the wasteful nature of free-market competition, using means of consolidation and monopoly. [14] Nevertheless, Rockefeller's endeavors led to significant economic benefits for American society. [15] Under Standard Oil, kerosene was easier to obtain, making resources more affordable for ordinary citizens, while increasing efficiency expanded energy access across the United States. [16] Rockefeller, later in his life, donated roughly 540 million dollars to education, medicine, and scientific research, and funded institutions which continue to benefit society today, such as the United Nations or the University of Chicago. [17] Consequentialist thinkers argue that the morality of an action is determined by its outcomes rather than the motives behind it; thus, they would argue that Rockefeller's positive contributions are what truly matter. This could be logical; after all, intentions exist solely in a person's mind, when outcome affects the real world. If a positive outcome is shared, the tangible benefit is realized regardless of the motive. However, this view is limited and has its flaws. It overlooks how harmful intentions may lead to manipulation and repetitive unethical behaviour. But if society judged actions solely by their results, a maelstrom of destruction would likely be unleashed. Individuals would feel persuaded to justify greed, manipulation, or exploitation, if they happened to produce beneficial outcomes. Furthermore, good consequences often turn out to be accidental or unpredictable, whereas intentions are carved from an individual's principles, and gives a glimpse at what type of person they are. This distinction emphasizes a limitation on ethics based on outcomes: though positive consequences deserve to be recognized and utilized, moral concerns arise when motives encourage unethical conduct and weaken public trust. Therefore, Kant's approach would ensure more certainty; a human's mind and motivations should correlate with what they choose to physically do in real life.

Trust, Character, Society

Ultimately, a favorable outcome cannot vindicate an action born from unethical motives; true morality requires both righteous ends and noble means. As Kantian deontology has demonstrated, morality depends not only upon what has been achieved, but also objectives that guide one's actions. Morality remains essential; it serves as the bridge between communities in society, letting people trust, betray, and cooperate with others as they wish. If bad intent, sculpted to form good outcomes, becomes normalized in modern society, people will feel less pressured to obey the law. This would also reinforce the idea that success is more important than integrity.

Human societies depend on the expectation that people will act according to ethical principles. This idea echoes the ideas of Jean de La Bruyère, who emphasized the importance of character when evaluating human behaviour. [18] A society that values moral intentions alongside positive outcomes would foster genuine trust and virtue, serving as the glue keeping communities together. Therefore, it can be wrong to do the right thing for the wrong reason.

Works Cited

The. "Characters, or the Manners of the Age, with the Characters of Theophrastus | Work by La Bruyère | Britannica." Encyclopedia Britannica, 2026, www.britannica.com/topic/Characters-or-the-Manners-of-the-Age-with-The-Characters-of-Theophrastus. .

Muhammad Wahaj. "Bridging Kantian Ethics and Consequentialism." Medium, 28 Aug. 2025, medium.com/@wahaj67/bridging-kantian-ethics-and-consequentialism-e3f9b6cb5d0a.

White, Mark D. "Getting Good Results vs Doing the Right Thing." Www.learnliberty.org, 17 Nov. 2016, www.learnliberty.org/blog/getting-good-results-vs-doing-the-right-thing/.

Alexander, Larry. "Deontological Ethics." Stanford Encyclopedia of Philosophy, 11 Dec. 2024, plato.stanford.edu/entries/ethics-deontological/.

Britannica. "Categorical Imperative." Encyclopædia Britannica, 27 Dec. 2017, www.britannica.com/topic/categorical-imperative.

Johnson, Robert, and Adam Cureton. "Kant's Moral Philosophy." Stanford Encyclopedia of Philosophy, Stanford University, 2022, plato.stanford.edu/entries/kant-moral/.

"Who Decides Right and Wrong: Where Morality Really Comes From." Rootedindecency.com, rootedindecency.com/blog/values/what-is-morality-definition.html.

Cantor, Alan. "The "Charity" of Elon Musk – Alan Cantor Consulting." Alan Cantor Consulting, 20 Mar. 2024, www.alancantorconsulting.com/2024/03/the-charity-of-elon-musk/.

"What Is a "Payout Requirement" for a Private Foundation?" Candid, 2026, learning.candid.org/private-foundation-payout-requirement/278315.

Fahrenthold, David A, and Theodore Schleifer. "Musk's Foundation Gave Away Less Money than Required in 2023." The New York Times, 12 Dec. 2024, www.nytimes.com/2024/12/12/us/politics/musk-foundation-taxes-donations.html.

"Lecture 11: Mind Your Motive – Harvard Justice." Justiceharvard.org, 28 Dec. 2016, justiceharvard.org/lecture-11-mind-your-motive/#1477505675034-57204795-9b61.

"The Calculating Shopkeeper … [IMMANUEL KANT (MICHAEL J. SANDEL)] | Part A'." Lectures Bureau, 5 Mar. 2022, www.lecturesbureau.gr/1/the-calculating-shopkeeper-part-a-1761a/?lang=en.

History.com Editors. "John D. Rockefeller - Biography, Facts & Children | HISTORY." HISTORY, 9 Apr. 2010, www.history.com/articles/john-d-rockefeller.

"Bulk Fuel Australia." Bulkfuel.com.au, 2018, bulkfuel.com.au/news/the-controversial-history-john-d-rockefeller.

Dwyer, James. "The Impact of Rockefeller." Medium, 22 Oct. 2019, medium.com/@jamesdwyer123456/the-impact-of-rockefeller-8fef37a21084.

"Burning a Monopoly: Kerosene and Anti-Monopoly Politics against Standard Oil, 1848-1911 | Consortium for History of Science, Technology and Medicine." Chstm.org, 2023, www.chstm.org/news/burning-monopoly-kerosene-and-anti-monopoly-politics-against-standard-oil-1848-1911.

"The Rockefellers: A Legacy of Giving | Rockefeller Philanthropy Advisors." Rockefeller Philanthropy Advisors, 2017, www.rockpa.org/guide/rockefellers-legacy-giving/.

"30 Best Jean de La Bruyere Quotes with Image | Bookey." Bookey.app, 2023, www.bookey.app/quote-author/jean-de-la-bruyere.

From Disparity to Equity: Understanding the Racial Wealth Gap and the Future of African American Wealth By Braeland Moten

Abstract

This paper examines the racial wealth gap in the United States by exploring how wealth is understood, how historical barriers have shaped economic opportunity, and how these disparities continue to affect African American families today. Beginning with definitions of wealth and a student survey on teenage wealth perceptions, the paper connects individual understandings of wealth to broader structural inequalities. It then analyzes historical barriers such as redlining, unequal wages, and educational inequality, as well as present-day challenges involving income, employment, credit access, investing, and financial literacy. Finally, the paper considers pathways toward greater wealth equity through education, economic access, community awareness, and financial knowledge. Overall, this paper argues that the racial wealth gap is not the result of individual choices alone, but of generations of interconnected barriers that must be addressed to expand opportunity for future generations of African Americans.

What does it truly mean to be wealthy? Is wealth defined by a mansion, a billion dollars, a large family, or something deeper? While wealth is often understood as money or material possessions, in reality it is shaped by much more than financial assets alone. Definitions of wealth vary across individuals and communities, often influenced by personal experience, upbringing, and access to opportunity. This paper examines how wealth is perceived and how those perceptions connect to a larger and more complex issue in the United States: the racial wealth gap. While personal definitions of wealth differ, the ability to build and maintain wealth is not equally distributed across all groups. Historical and ongoing barriers in education, employment, housing, and financial resources have contributed to long-standing disparities in wealth building in African American communities. The purpose of this paper is to present the causes and consequences of the racial wealth gap while exploring pathways toward greater wealth equity for future generations of African Americans.

To explore what wealth really means, I consulted several sources: general definitions from Wikipedia, teenage wealth perceptions from Junior Achievement, and a statistical study on financial literacy from Science Direct. Although definitions of wealth vary by source, it is generally described as the abundance of valuable financial assets or physical possessions that can be converted into currency—including houses, cars, cash, and investments. This makes it easy to simply associate wealth with having lots of money, but the reality is more complex than that.

One's definition of wealth is affected not only by financial factors but also by social and cultural factors, which contribute to how individuals, specifically teens, perceive wealth. For example, teenagers raised in financially stable households often have a clearer idea of what "wealth" is, likely associating it with investments and security. However, for those who grow up in lower-income households or communities with fewer resources, wealth may be more immediate things such as the ability to provide for their families or have stable income. Research

from The British Journal of Social Psychology on social class and socioeconomic status suggests that people's perception of wealth is heavily influenced by the material conditions they grow up in, as well as by their sense of social rank and identity. Ultimately, a person's understanding of wealth reflects not only their financial situation but also the social environment and life experiences that shape how they see economic success.

To get an idea of what my teenage peers thought about wealth, I collected information from a voluntary survey of students at my high school. Students were asked to give their definition of wealth in one sentence, express what they think makes somebody wealthy, and identify one thing they would own to be considered wealthy—including a few background questions. One of those questions was, "If you could own just ONE thing to be considered wealthy, what would it be?" This survey featured 21 responses from students in an 11th grade American Literature class.

Many responses reflected material definitions of wealth. For question three, "If you could own just ONE thing to be considered wealthy," 15 out of 21 responses included monetary items, most commonly cash, a large stock portfolio, or a mansion. A few responses also expressed wealth as having a nice family, love, and happiness. From what we have discussed so far, it should come as no surprise that wealth is viewed as mainly material, and there is nothing necessarily wrong about that; however, what was more intriguing about the survey was each definition of wealth itself. In response to the question "In one sentence how would you define wealth?" respondents were generally less materialistic, with only 7 of 21 answers specifically stating the definition in relation to money or assets but rather as financial security, strong relationships, happiness, and time. From this survey, the pattern suggests that teenagers often associate wealth with visible possessions; however, many also understand it more broadly as stability and security. One of the most important factors in shaping that understanding is financial literacy education. According to the National Association of State Boards of Education, 41 states now require some form of financial literacy course for high school graduation. While it isn't all 50 states, it is an important step toward ensuring all students have a foundation for understanding wealth.

As we have discussed, teens think about wealth in different ways. Family, friends, education, and cultural background all play a key role in how a teen defines wealth. Ultimately, we must all understand that wealth is more than just money. These many factors will influence how today's young people approach saving, spending, investing, and building wealth throughout adulthood. Understanding wealth early is therefore not simply a matter of personal finance; it is an important first step toward creating greater financial opportunity for future generations.

While individual perspectives help us understand what wealth means, they do not explain why wealth is distributed so differently across the United States. The ability to build wealth is shaped not only by personal choices but also by the opportunities available to individuals and families over generations. For many African American families, those opportunities have been limited by various historical barriers that have restricted access to education, homeownership, employment, and other pathways to economic opportunity. Understanding these barriers and the

result of them is essential to understanding not only the origins of the racial wealth gap but also why it persists today.

According to the Federal Reserve, White households account for 83.5% of the nation's wealth, compared to roughly 3.3% held by Black households. This is not the result of individual financial decisions alone but historical barriers that made it difficult for wealth to be built in Black communities over time. These barriers, including redlining, inequities in employment, and unequal access to quality education, have all led to limited opportunities for Black families to build and pass down generational wealth.

Redlining was one of the most powerful barriers that prevented Black families from building wealth through homeownership. Its effects are still seen today, as formerly “redlined” neighborhoods remain lower income and predominantly minority areas. Amidst the height of the Great Depression in 1933, President Franklin D. Roosevelt started the Home Owners' Loan Corporation (HOLC) in an attempt to make purchasing homes more affordable for first-time home owners. However, when the HOLC began to disburse mortgages, it developed a scheme that helped them to decide which neighborhoods and communities were deserving of loans and which ones were not, called redlining. They made maps; green areas marked a community you could lend to and red areas you could not. These red areas consisted almost entirely of people of color, making it nearly impossible for them to obtain a home loan. As a result, home ownership was rarely an option for minorities. Redlining lasted until 1968, when the Fair Housing Act was passed. However, irreversible damage had already been done to Black communities. The lack of homeownership, a key to long-term financial stability, and decades without the opportunity to reap the benefits of home equity growth stunted the ability for wealth to be built, widening the racial wealth gap even more.

This historical wealth gap is also rooted in long-standing unequal wages and inequities in employment opportunities. After slavery was abolished in 1865, African Americans entered a labor system that continued to restrict economic opportunity through low wages and limited employment options. Black people worked mainly as day laborers on farmland for low wages. Although this was an extreme step up from slavery itself, it established a pattern that would continue for generations, in which African Americans would be paid far less than White Americans. This wage inequality would persist through the mid-20th century. According to the U.S. Bureau of Labor Statistics, in 1979 Black men earned roughly 80% of what White men earned, a figure that surprisingly declined in the following decades to 70% as of 2016. Although small improvements were made, especially through the civil rights movement from 1954 to 1968, discrimination and inequality would continue. Low wages and the struggle for Black Americans to find work resulted in the lack of stable and sufficient income. Therefore, Black households were unable to build a strong foundation for long-term wealth.

Beyond employment and wages, unequal access to education created another major barrier that limited economic opportunity for African Americans. Education plays a critical role in determining future career opportunities, income potential, and the ability to build wealth over time. Since the arrival of the first Africans in what is now Virginia, Black people have gone

without quality education, or education entirely. From 1740 to 1834, many southern states passed anti-literacy laws, making it illegal to teach enslaved and free Black people to read or write. Although some were educated by free Blacks, religious figures, white abolitionists, or even themselves, a large majority remained illiterate. Years later, in 1865, with the passing of the 13th amendment, anti-literacy laws came to an end, and education was now “accessible” by all. However, educational opportunities for African Americans remained far from equal, as segregation within schools would last until 1954. In the time leading up to desegregation, Black students attended underfunded schools with fewer resources, larger class sizes, and limited access to quality materials and experienced teachers. Consequently, Black students simply did not receive the same level of academic preparation as their White counterparts, limiting opportunities for higher education and well-paid careers.

Generational wealth, or financial assets passed down through generations, is one of the most powerful factors enabling economic opportunity, and it explains why the gaps discussed in this article continue across decades. When families are able to accumulate assets such as homes, investments, and savings, the next generation starts in a stronger situation than the one before it. However, because of these many historical barriers such as discriminatory housing policies, unequal wages, and inequity in education, Black families are decades behind in developing generational wealth.

In addition to historical barriers, there are many economic challenges that African Americans face today that are not a coincidence, but the result of the United States' own systems. The racial wealth gap has existed throughout much of American history and continues today. According to Lending Tree, as of 2024, Black households earned an average of $56,020 compared to $84,630 for White households, a 36.3% gap. This is the result of various economic challenges African Americans face today, most notably, the income gap itself, employment barriers, unequal access to credit and loans, and disparities in investing and spending.

One of the most significant factors contributing to the racial wealth gap is the persistent income gap, rooted in a long history of undervaluing Black labor. The current gap in yearly household income is at a staggering 36.3%. Lower household income makes it more difficult for families to afford essential expenses such as housing, transportation, insurance, and childcare. In addition, rising costs for necessities such as groceries, fuel, and childcare place an even greater financial strain on Black households, leaving less income available for saving and investing. As a result, many Black families have fewer opportunities to save, invest, and accumulate assets over time, making it far more difficult to build and pass down wealth.

Employment barriers are also a key component in today's economic disparities, limiting access to jobs and long-term career advancement for African Americans. Black workers face persistent obstacles in obtaining stable, well-paying jobs, mainly hiring discrimination and unequal access to quality positions across industries. Consequently, building wealth becomes much harder, as the focus is more on getting by rather than investment and saving. On the note of unemployment itself, according to the Bureau of Labor Statistics as of 2026, African Americans ages 16 and above have an unemployment rate of 7.4%, the largest between Whites, Asians, and

Hispanics. This disparity in unemployment is not by chance, but part of a broader pattern that reflects an unfair labor system. According to the Center for American Progress, African American workers experience what is called “last hired, first fired," in which Black workers find themselves being the first ones dismissed from work due to economic downturn and being the last ones hired when the economy stabilizes again. The same report also notes that African Americans also face occupational segregation, which pushes them into lower-paying jobs with fewer benefits and less stability. Without a “sufficient” job, the ability for African Americans to accumulate wealth is difficult and over the last decades has contributed greatly to the racial wealth gap.

Access to credit and loans are also a major factor in wealth building, but Black Americans face barriers that make borrowing more difficult and expensive, making financial growth less likely. Credit is not just about borrowing money; it’s one of the main ways families buy homes, start businesses, and handle emergencies. The main barrier that affects African Americans in relation to credit is predatory lending. Research from the NAACP and The Leadership Conference Education Fund tells us that people of color are 2 to 3 times more likely to be subject to predatory loans. Predatory loans often have high interest rates, hidden fees, and confusing terms that mislead borrowers, leading them to take out more loans just to keep up with payments. These loans can create a cycle of debt instead of helping build wealth. Historically, mortgage redlining and similar lending practices have long denied communities of color the chance to build generational wealth through homeownership. Although it is not as applicable to the general African American population, issues with small business lending also create an anchor for asset building. According to a study by CFPB, Black small business applicants received less favorable treatment than white applicants in matched-pair testing. The struggle to obtain small business loans can affect both those trying to open a business and those already running one; they either do not have the opportunity for profit through business or have to close a business due to lack of needed capital. Access to personal loans, credit, and fair lending is key to building wealth, and when Black Americans are denied fair access to these resources, it becomes much harder to buy homes, start businesses, and create long-term financial stability. Instead of helping families grow, unequal lending can trap them in debt and keep wealth from accumulating across generations.

The many factors we have discussed, ranging from actual income, employment, and loans, all play into the most major factor in building wealth – money. Money is what is needed to build more money, mainly through investing. It is estimated that between 50 and 60 percent of millionaires were made through long-term, strategic investing. But what does it take to invest? Capital. However, capital is only half the battle; earning income is enough of a struggle for millions of African Americans, but having the resources to turn that income into lasting wealth is another major barrier. According to Pew Research Center, White Americans are nearly twice as likely to own stocks as Black Americans. The reason for this is most likely a result of the factors we have already discussed: income, employment, and credit. The gap in these things between Whites and Blacks has caused a lack of opportunity to invest and accumulate assets over

generations, which is how the bulk of wealth is built in investments, long-term. The lack of ability for investments to be made is a large contributor to the wealth gap, and only when we begin to close the gaps in other areas, such as income and employment, can we enable these other wealth-building opportunities like investing.

Again, the racial wealth gap in African American communities is not the result of one single issue but generations of interconnected barriers. Understanding these barriers is essential, but reducing the wealth gap will require addressing the systems that continue to limit opportunity today. Among the most significant of these systems is education, where unequal access to resources and opportunities continues to shape financial outcomes long after students leave the classroom.

Education is often presented as the key to success; however, not every student has equal access to quality schools, financial education, or the resources needed to prepare for the future. Educational and opportunity gaps continue to shape the experiences of African American students in ways that extend beyond the classroom. These disparities can influence academic achievement, postsecondary pathways, and long-term economic outcomes. These gaps are shaped by differences in school resources, academic opportunities, and support systems that influence students' ability to succeed after graduation. In the classroom, research from the School Finance Indicator Database shows African American students are more likely to attend schools that have less funding, fewer college preparatory classes, and advanced learning opportunities, among other things, which limit their opportunities after high school.

Education inequality is deeply rooted in American history. Following the end of the Civil War, many states enacted laws that separated Black and White students into different schools under the doctrine of "separate but equal." These schools serving African American students had significantly fewer resources, funding, and facilities compared to White schools. These segregated schools existed for decades until the 1954 Supreme Court case Brown v. Board of Education declared school segregation unconstitutional. However, while segregation was no longer legal, the inequalities that this period created continue to influence education in African American communities.

Today, one of the clearest examples of educational inequality is reflected in the disparities in funding and resources available to schools that predominantly serve African American students across the United States. Public school funding is heavily dependent on local property taxes, meaning wealthier communities can generate more revenue for their schools, and lower-income communities provide less. Many predominantly African American communities, which on average have lower household income, generate less funding for local public schools. This causes the lack of many resources and opportunities within these schools. The lack of said resources diminishes the level of education provided to the attending students. School funding affects nearly every aspect of a student's educational experience. Teacher salaries, classroom materials, technology, extracurricular activities, and the offering of advanced courses are just a few things that are affected by low funding. These funding disparities often create a domino effect; schools may struggle to attract and retain experienced teachers, and students may have

less postsecondary preparation support due to the lack of counselors, other support staff, and advanced coursework opportunities. As a result of this educational inequality, students may enter adulthood being less prepared for further education and professional opportunities.

An education that includes financial literacy is an important tool that helps students learn the fundamental concepts of personal finance, such as budgeting and investing, preparing them to make informed decisions and plan for their financial future. For many students, the foundation of financial literacy is developed in the classroom through personal finance classes and other related courses. However, schools that predominantly serve African American students are less likely to offer personal finance courses due to the educational inequalities we have already discussed. Without a basic understanding of personal finance, young adults leave their high schools less prepared to manage the financial responsibilities of being an adult. This gap in financial literacy education is reflected in the P-Fin 8 index, which found that on average African American adults answered 38% of general personal finance questions correctly, compared to 55% of White adults.

The barriers discussed do not end when a student leaves the classroom but continue to follow them as they pursue higher education, enter the workforce, and attempt to build wealth. African American students with fewer educational resources will have fewer opportunities to prepare for college and careers, limiting their long-term opportunity to build wealth. For the estimated 36% of African Americans who pursue undergraduate education, the barriers do not end there. According to the National Center for Education Statistics, only 40% of Black students graduate with their bachelor's degree within 6 years. In addition, many graduates also leave with substantial student loan debt. According to the Education Data Initiative, 48% of Black borrowers reported having at least $25,000 of debt from their education. While further education can benefit those in search of employment, high levels of student debt can delay saving, investing, homeownership, and other important milestones for building wealth.

We can conclude that educational inequalities affect far more than a student's experience in the classroom. Differences in access to quality education, resources, and preparation can continue to shape outcomes well beyond high school. Ensuring more equitable access to educational opportunities remains an important step toward improving long-term outcomes for students. Understanding these barriers is only the first step. The more important question is how we begin to reduce them and expand opportunities for future generations. Because the racial wealth gap is the result of generations of historical, economic, and social barriers, no single law, program, or individual action can eliminate it entirely.

Education serves as the foundation upon which long-term economic opportunity and financial stability are built. Expanding access to quality education for African American students is one of the most effective long-term investments that can be made toward wealth equity, providing knowledge and skills to make informed decisions throughout life. Supporting students extends beyond traditional academics by providing pathways toward higher education, careers, financial decision-making, and lifelong opportunity. As more African American students gain

access to these educational opportunities, more will be prepared to pursue higher education, enter skilled careers, start businesses, manage money effectively, and build wealth.

While education lays the foundation, lasting wealth equity also depends on eliminating economic barriers for African Americans through adulthood. Throughout this paper, we have explored the many barriers that continue to limit wealth-building opportunities for African Americans, including barriers in employment, wages, access to credit, entrepreneurship, and homeownership. Without steady income, it becomes much harder to save, invest, and allocate resources toward long-term wealth building. Addressing these challenges will require continued efforts from employers, financial institutions, policymakers, and communities to expand access to economic opportunities. Without continued progress in removing these barriers, wealth-building opportunities for future generations of African Americans will remain limited.

Although large-scale systemic change is essential to addressing the racial wealth gap, meaningful progress can also be made through increased awareness, education, and action within African American communities. One of the first steps toward change is acknowledging that the racial wealth gap is not just a set of statistics but a real issue that affects access to opportunity, financial stability, and wealth building for African American families. Open conversations about these disparities can help shift how individuals understand wealth and highlight the importance of financial knowledge in shaping outcomes for future generations. Increasing awareness of the racial wealth gap can encourage individuals, families, and communities to better understand the barriers that affect wealth building and seek out resources that support financial growth. Financial knowledge has never been more accessible. Schools, libraries, nonprofit organizations, community leaders, and countless free online resources now provide opportunities for individuals to learn about budgeting, saving, and investing. Engaging with these tools, seeking out financial learning opportunities, and learning from educators or local community leaders can help individuals develop strong financial literacy. While these actions alone cannot eliminate systemic inequality, they can equip the next generation with knowledge that improves their ability to make informed financial decisions and build wealth in their adult lives.

Ultimately, wealth is about far more than financial assets or material possessions; it represents security, opportunity, and the ability to create a better future. Throughout this paper, it has become clear that the racial wealth gap is not the result of one individual decision or one isolated issue but generations of interconnected historical, economic, and social barriers that have shaped the opportunities available to African American families. Understanding these barriers is an important first step toward creating meaningful change. The future of wealth equity depends on the opportunities that are created for the generations to follow. The challenges that have limited African Americans' ability to build wealth for centuries should not determine the economic possibilities available to future generations. By expanding access to quality education, increasing financial knowledge, and creating fairer economic systems, we can help ensure that future generations are better prepared to build, preserve, and pass down wealth. The goal of wealth equity is not simply to eliminate a statistical difference between groups, but to create a society where individuals have a fair opportunity to achieve financial security and provide for

their families. Although closing the racial wealth gap will require continued effort from communities, institutions, and policymakers, the actions taken today will influence the opportunities available tomorrow. The path toward wealth equity will be long and challenging, but creating a future where African Americans have the ability to build lasting wealth is worth pursuing.

Works Cited

American Bar Association. “Predatory Lending: The New Face of Economic Injustice.” American Bar Association, https://www.americanbar.org/groups/crsj/resources/human-rights/archive/predatory-lending-new-face-economic-injustice/.

Bureau of Labor Statistics. “Employment Status of the Civilian Population by Race, Sex, and Age.” U.S. Bureau of Labor Statistics, https://www.bls.gov/web/empsit/cpsee_e16.htm.

Bureau of Labor Statistics. Labor Force Characteristics by Race and Ethnicity, 2023. U.S. Bureau of Labor Statistics, https://www.bls.gov/opub/reports/race-and-ethnicity/2023/.

Bureau of Labor Statistics. “The Unexplainable Growing Black-White Wage Gap.” Monthly Labor Review, U.S. Bureau of Labor Statistics, https://www.bls.gov/opub/mlr/2017/beyond-bls/the-unexplainable-growing-black-white-wage-gap.htm.

Consumer Financial Protection Bureau. Consumer Financial Protection Bureau, https://www.consumerfinance.gov/.

Federal Reserve Board. “Distribution of Household Wealth in the U.S. since 1989.” Federal Reserve, https://www.federalreserve.gov/releases/z1/dataviz/dfa/distribute/table/.

Franklin, Charles. “What Is Wealth?” Goucher College, https://www.gbc.edu/about/news/what-exactly-is-wealth/.

Junior Achievement USA. “Money Without Borders Survey: Only 27% of Teens Think National Governments Have Most Influence on Global Economy.” Junior Achievement USA, https://centralva.ja.org/news/press-releases/-money-without-borders-survey-only-27-of-teens-think-national-governments-have-most-influence-on-global-economy.

Manstead, Antony S. R. “The Psychology of Social Class: How Socioeconomic Status Impacts Thought, Feelings, and Behaviour.” British Journal of Social Psychology, vol. 57, no. 2, 2018, pp. 267–291. Wiley Online Library, https://doi.org/10.1111/bjso.12251.

National Archives. “13th Amendment to the U.S. Constitution: Abolition of Slavery.” National Archives, https://www.archives.gov/founding-docs/amendments-11-27#xiii.

National Archives. “Brown v. Board of Education.” National Archives, https://www.archives.gov/milestone-documents/brown-v-board-of-education.

National Association of State Boards of Education. “Advancing High Schoolers’ Financial Literacy.” NASBE, https://www.nasbe.org/advancing-high-schoolers-financial-literacy/.

National Center for Education Statistics. “Status and Trends in the Education of Racial and Ethnic Groups.” NCES, https://nces.ed.gov/programs/raceindicators/indicator_red.asp.

National Community Reinvestment Coalition. “CFPB Study Confirms Advocates’ Findings That Small Business Lending Has a Big Racial Discrimination Problem.” NCRC, https://ncrc.org/cfpb-study-confirms-advocates-findings-that-small-business-lending-has-a-big-racial-discrimination-problem/.

National Museum of African American History and Culture. “The Struggle Against Segregated Education.” Smithsonian National Museum of African American History and Culture, https://nmaahc.si.edu/explore/stories/struggle-against-segregated-education.
National Urban League. “Disparities in African American Financial Literacy.” National Urban League, https://nul.org/news/disparities-african-american-financial-literacy-serious-problems-concept.
Pew Research Center. “A Booming U.S. Stock Market Doesn’t Benefit All Racial and Ethnic Groups Equally.” Pew Research Center, 6 Mar. 2024, https://www.pewresearch.org/short-reads/2024/03/06/a-booming-us-stock-market-doesnt-benefit-all-racial-and-ethnic-groups-equally/.
San Diego Foundation. “What Is Generational Wealth?” San Diego Foundation, https://www.sdfoundation.org/news-events/sdf-news/what-is-generational-wealth/.
School Finance Indicator Database. School Finance Indicator Database Annual Report 2024. https://www.schoolfinancedata.org/sites/default/files/data/sfid2024_annualreport.pdf.
ScienceDirect. ScienceDirect, https://www.sciencedirect.com/.
The Annie E. Casey Foundation. “Racial Inequality in Education.” The Annie E. Casey Foundation, https://www.aecf.org/blog/racial-inequality-in-education.
The Global Financial Literacy Excellence Center. “P-Fin Index.” GFLEC, https://gflec.org/initiatives/pfin8-index/.
The Leadership Conference Education Fund. “These Are the Significant Costs of Predatory Loans.” The Leadership Conference Education Fund, https://civilrights.org/edfund/resource/these-are-the-significant-costs-of-predatory-loans/.
United States Department of Justice. “Fair Housing Act.” Department of Justice, https://www.justice.gov/crt/fair-housing-act-1.
United Way. “Racial Inequality in Education.” United Way, https://unitedwaynca.org/blog/racial-inequality-in-education/.

Mutation Directed Therapies in Pediatric and Adult Gliomas By Ekaterine Maghlakelidze and Mikheili Sharvashidze

Abstract

Glioblastoma Multiforme (GBM) remains one of the most aggressive and treatment-resistant forms of cancer. Standard treatments like chemotherapy, radiotherapy, and surgery are often futile due to the tumor infiltrating critical parts of the brain and severe treatment-related toxicities (Johns Hopkins Medicine, 2022). However, advancements in the medical field have enabled mutation-specific targeted therapies, which increase the likelihood of survival and quality of life during treatment. This retrospective literature review evaluates mutation-directed therapies in pediatric and adult gliomas, with a primary focus on IDH1, BRAF V600E, and NF1-mutated tumors. Data were analyzed from clinical trials that investigate treatments such as Ivosidenib, Tovorafenib, Dabrafenib ± Trametinib, everolimus, neoadjuvant/adjuvant vaccines, and the oncolytic virus DNX-2401. Outcomes of interest included progression-free survival (PFS), overall response rate (ORR), safety profile, and, if applicable, the comparison between the available treatments. Overall, the studies demonstrated significant improvements in disease control and progression-free survival while generally maintaining favorable safety profiles. For pediatric patients suffering from low-grade gliomas with BRAF alterations, Tovorafenib achieved an ORR of 67%, and the combination therapy of Trametinib + dabrafenib managed to triple the progression-free survival, compared to the data accumulated from chemotherapy. DNX-2401, combined with radiotherapy in diffuse intrinsic pontine glioma, showed disease control and tumor shrinkage in 75% of the patients, managing to also improve median survival to rates higher than ever estimated. Everolimus provided 71.1% of patients with 2 years of PFS, and in cases of non-enhancing mIDH1 adult gliomas, the drug Ivosidenib brought the median PFS up to a staggering 36.8 months. Limitations included short follow-up times, small sample sizes, and the diversity of the patient profiles. Mutation-based therapies display superior results with not only improved survival and progression rates but also a more promising safety profile, which is directly related to the patient's health and advancements.

1. Introduction

Glioblastoma (GBM), also referred to as glioblastoma multiforme, is the most common and aggressive type of cancer originating from the brain. It is caused by DNA mutations, resulting in uncontrollable cell growth, and it is thought that glioblastomas carry more genetic abnormalities. GBMs are most commonly known for their rapid progression, infiltrative nature, and resistance to treatment. This form of cancer is exceptionally difficult to treat due to its location, making it, most of the time, inoperable. It infiltrates healthy tissue and impacts critical brain regions. Chemotherapy is also accompanied by severe treatment-related adverse events that make it a less than ideal course of treatment for patients suffering from a brain tumor that is also classified as a grade IV astrocytoma by the World Health Organization. This treatment is

generally more effective when combined with other types of therapies, such as radiation or medication. Due to all of this, life expectancy with treatment ranges anywhere from 12-18 months, and a staggering 3-4 months if the aggressive cancer is not addressed. Patients with glioblastoma deal with severe symptoms such as headaches, seizures, alterations in mental functions, memory loss, and much more. Some of the mutations that cause this disease include IDH1, BRAF V600E, and Neurofibromatosis type 1(NF1). This paper will address alternative treatments that may offer more promising results with fewer severe side effects and provide hope for future research. These therapies include Ivosidenib for mIDH1, Tovorafenib, a type II RAF inhibitor, Dabrafenib, everolimus, vaccinations, and a tumor-irradiating virus DNX-2401.

2. Methodology

All of the following information will be cited and analyzed from the searchable database, PubMed. Clinical trials were chosen based off of the timeline, since they had to have been written in the last 5 years, and they had to concern either a pediatric or adult population, which suffered from gliomas. The texts were analyzed based on their end results and primary points. Evaluating whether or not they were able to produce a breakthrough or discover something that could possibly impact future research. The discussion is divided into two parts. One concerns solely adult gliomas, and the other pediatric patients. This is done to make the information more easily obtainable and further highlight the importance of individualized, specific therapies.

3. Clinical trial summaries

One clinical trial assessed the efficacy and safety of the drug Ivosidenib in Isocitrate Dehydrogenase-mutated advanced cases of Gliomas (Mellinghoff et al., 2020). IDH1 mutations are present in over 70% of lower-grade gliomas (LGGs), which, despite their slow progression, eventually turn into more aggressive types of tumors. The trial enrolled 66 patients, 35 with non-enhancing gliomas and 31 with enhancing gliomas. The patients only qualified if they were aged 18 or older, with no radiation treatments within six months of enrollment. The study was divided into two phases: a dose-escalation phase, where patients received daily oral doses ranging from 100 mg to 1200 mg, and a dose-expansion phase, which continued with the optimal dose of 500mg. No dose-limiting toxicities were observed, and the maximum tolerated dose was not reached. In order to monitor and accurately check the effectiveness of this particular drug, patients were required to undergo at least three MRI scans, each spaced a minimum of two months apart, allowing investigators to assess tumor changes over time. These MRIs, compared to the ones after starting the drug trial, revealed tumor stabilization and even shrinkage in some cases. Before initiating Ivosidenib, the six-month progression rate among patients was approximately 25%; this dropped significantly to just 9% during treatment. Progression-free survival (PFS) also significantly improved; the median PFS across all patients was 13.6 months, but when broken down by tumor type, those with non-enhancing gliomas had a notably longer median PFS of 36.8 months compared to just 3.6 months in patients with enhancing gliomas. The drug had a larger success rate with non-enhancing gliomas, most likely because they tend to

be in the earlier stages with fewer genetic mutations. Adverse events (AEs) were mostly low grade and did not hinder the process, with the most common being diarrhea, nausea, fatigue, seizures, and elevated liver enzymes. Overall, the trial demonstrated that Ivosidenib is well tolerated and may offer substantial clinical benefit, particularly in patients with non-enhancing mIDH1 LGGs.

A clinical trial conducted by Kilburn et al. (2024) is a phase 2 study investigating the efficacy and safety of the oral drug Tovorafenib, specifically in relapsed / refractory LGGs. Tovorafenib is an oral CNS-penetrant type II RAF inhibitor. This clinical trial selected pediatric patients under the age of 25 with BRAF-altered progressive low-grade gliomas (pLGGs). In total, the study was conducted with 77 patients, all of whom had undergone at least one systematic therapy in their lifetime. The primary endpoint is overall response rate, while the secondary endpoint includes duration of response, as well as safety. The results were extremely promising, with an ORR of 67%, with one complete and 51 partial responses. The disease control rate reached a staggering 93%. Median PFS was 16.6 months, and median DOR was 16.7 months, indicating both substantial and sustained efficacy. In terms of safety, most of the adverse events were grade 1 or 2, which demonstrates a favorable safety profile. Alopecia was most common (58%), followed by fatigue and skin effects. Only 17% of patients experienced grade 3 or higher treatment-related events, and no treatment-related deaths occurred. These findings suggest that Tovorafenib may be a highly effective and well-tolerated targeted therapy for children and young adults with BRAF-altered pLGG, representing a significant advancement in pediatric neuro-oncology.

Dabrafenib plus Trametinib in Pediatric Glioma with BRAF V600 Mutations (Bouffet et al. 2023) is another phase 2 clinical trial that discusses the safety and effectiveness of combining dabrafenib and trametinib, in contrast to standard chemotherapy. The eligible patients were children aged 1-17 years old, with newly diagnosed BRAF V600 harboring LGGs. The patients were assigned in a random 2:1 ratio to receive either the standard or the targeted therapy. The main goal was oriented towards the overall response rate, which demonstrated great promise. Nearly 50% of the patients receiving dabrafenib + trametinib displayed tumor shrinkage/reduction and positively responded to the medication. This shows great improvement, compared to the 10% yielded from chemotherapy. Additionally, the PFS was three times longer, and the children receiving targeted therapy maintained disease control for over 6 months. The safety profile was also more favorable, with fewer patients experiencing severe treatment-related side effects. Overall, these results indicate much promise that the combination therapy entailing dabrafenib and trametinib yields better results and superior clinical benefits in contrast to the standard chemotherapy, when it comes to pediatric patients with BRAF V600 mutant LGGs.

The study “Trametinib in Pediatric Patients With Neurofibromatosis Type 1–Associated Plexiform Neurofibroma or Low-Grade Glioma: A Multi-Part, Phase I/II Trial” (Hargrave et al. 2023) is a four part clinical trial that explores the pediatric dosage, as well as safety, of Trametinib with or without Dabrafenib and a disease specific cohort that touches on LGGs stemming from the BRAF V600 mutation. The patients were picked based on their age (<18),

and they had to have relapsed / refractory tumors. The trial was divided into a few parts: Part A focused on Trametinib monotherapy dosage, and Part B concerned disease-specific expansion. Dabrafenib + Trametinib dose finding and disease-specific expansion were discussed in parts C and D, respectively. The study employed a total of 139 pediatric patients. 91 were treated with monotherapy, and the other 48 with combination therapy. 49 of those patients had BRAF V600-mutant glioma, including 47 with LGGs and 2 with HGGs. The primary goal was to establish pediatric-appropriate dosing and to determine whether the combination therapy offered improved outcomes. For trametinib, the recommended phase II dose (RP2D) was defined as 0.032 mg/kg once daily for patients under 6 years of age, and 0.025 mg/kg once daily for those aged 6 years and older. These doses successfully achieved steady-state pharmacokinetic targets in both monotherapy and combination arms. Dabrafenib was dosed according to prior established pediatric regimens. The safety profile was manageable, though dose-limiting toxicities (DLTs) were noted in the monotherapy dose-finding stage (Part A), including mucosal inflammation and hyponatremia. No DLTs were observed in the combination therapy dose-finding stage (Part C), suggesting that the combination regimen was generally well tolerated. Notably, treatment discontinuation due to adverse events was more common in the monotherapy group (54%) than in the combination group (22%). The overall response rate (ORR) was 15% in the trametinib monotherapy group, and 25% in the combination group. These results demonstrate clinically meaningful antitumor results, particularly in the combination arm. Progression-free survival (PFS) showed promise , with longer disease control observed in patients receiving both trametinib and dabrafenib. Overall, the study concluded that trametinib, especially when combined with dabrafenib, offers a favorable risk-benefit profile and represents a viable targeted therapy option for pediatric patients with BRAF V600–mutant low-grade gliomas.

The clinical study “ Neoadjuvant Vaccinations in Low-Grade Glioma” (Pellegatta et al. 2021), evaluated the safety and immune effect of a tumor cell-lysate vaccine that was derived from a Glioma stem cell line, referred to as GBM6-AD. These cells were harvested from a patient with glioblastoma multiforme. GBM6-AD expresses several immunogenic antigens that can target LGG cells and protect HGGs. This clinical trial was divided into two arms. Arm 1 received the said vaccines both before and after surgery, meaning neoadjuvant + adjuvant. On the other hand, patients involved in arm 2 received adjuvant-only vaccines. This study took place between 2016 and 2019. A total of 33 patients were scanned, but ultimately only 17 were included in the final analysis. All of the patients involved had an IDH1 mutation. In terms of safety, the vaccine was not associated with any serious AEs or dose-limiting toxicities. The median PFS was 11.0 months, while EFS was 23.7 months. No significant differences were observed in these results between the two arms. However, immune testing did show that the patient's arm 1 developed more advanced immune signals, as well as CD8+ T cells, which were trained to attack tumor cells. In summary, administering the vaccine in the neoadjuvant setting was safe and even resulted in a better immune response, although it did not lead to clear indications of how long the tumors remained dormant, these results show the need for further

research on this topic.

"A phase II study of continuous oral mTOR inhibitor everolimus for recurrent, radiographic-progressive neurofibromatosis type 1–associated pediatric low-grade glioma: a Neurofibromatosis Clinical Trials Consortium study"(Ullrich et al. 2020) is a phase II clinical study that evaluated the use of an oral drug, everolimus, that blocks mTOR pathways, a protein that plays a crucial role in the growth and survival of cells. This clinical trial specifically concerned children and young adults with neurofibromatosis type 1 and LGGs that had either come back or worsened after previous treatment. The mTOR pathway is recognized to have been overactive in NF1; therefore, a drug that can successfully stop it can play a crucial role in the shrinkage of tumors. A total of 23 patients aged between 3.2 and 21.6 years were enrolled. Said patients took everolimus by mouth for 48 weeks. The dose was completely dependent upon body surface area, and treatment continued unless any serious side effects occurred. Although the oral drug was deemed safe and generally well tolerated. Most side effects were mild, such as mouth sores and low blood count. While some patients were more affected by neutropenia and elevated liver enzymes. However, it should be noted that none of the patients had to stop treatment due to side effects. Out of the 22 patients that could be evaluated, one had a complete response, two had partial responses, and twelve had stable disease. Only seven patients showed tumor progression by the end of the 48 weeks. In terms of time-based outcomes, the 6-month progression-free survival (PFS) rate was 93.5%, meaning most patients had no tumor growth for at least six months. The 1-year PFS was 84.4%, and the 2-year PFS was 71.1%. This suggests that everolimus provided durable disease control in many cases. The event-free survival (EFS) was also promising, with a 2-year EFS rate of 61.1%, reflecting how long patients went without. Overall, 68% of patients had either a reduction in tumor size or stable disease. At the time of the analysis, all of the patients were alive, with a follow-up time of about 33 months. After testing using specific markers, it was also proven on a biological level that the mTOR pathways were being blocked. However, the drug level in the patient's blood varied, causing the different results, most likely due to individual metabolism and absorption. In summary, the oral drug everolimus showed extremely promising results by successfully stopping the overreaction of the mTOR pathways in patients suffering from NF1 and LGGs, who had previously had no luck with chemotherapy.

This early phase clinical trial explored the potential benefits and safety of injecting a tumor killing virus called DNX-2401 in pediatric patients suffering from diffuse intrinsic pontine glioma (DIPG). DIPG is a highly aggressive tumor, stemming from the brainstem, with minimal treatment options and a life expectancy of less than a year after diagnosis. The primary goal and ambition for this study was to directly inject the virus into the tumor, hoping to trigger an immune response that would ultimately attack the tumor cells and slow down the growth or even shrink it. There were 12 children selected ages ranging from 3-18 years old at a single centre. The first four received a lower dosage compared to the other eight, and eleven of these twelve patients went on to receive radiotherapy. Common side effects included headaches, nausea, and fatigue. Despite the side effects being mild and well tolerated, one patient developed

hemiparesis, weakness on one side, and another on all four limbs, Tetraperesis. After a median follow-up of 17.8 months, MRI scans were conducted, which showed tumor shrinkage in three patients, while eight had stable disease. Meaning that nine out of twelve children experienced a reduction or stabilization in tumor size. The median overall survival was 17.8 months, and two patients were still alive at the time of reporting, one showing absolutely no tumor progression at 38 months. Laboratory analyses displayed that DNX-2401 changed the microenvironment, enabling the production of T cells, therefore also triggering the immune response. In summary, injecting DNX-2401 directly into DIPG tumors in children, followed by radiotherapy, represented a significant chance of tumor stabilization as well as an acceptable safety profile. Most patients showed either tumor shrinkage or stabilization, and survival outcomes appeared better than historically expected for DIPG. While this data is from a small group, it supports further studies of oncolytic virus therapy in DIPG, and raises hope for new treatment strategies for this disease.

4. Discussion

4.1 Overview of Mutation-Directed Therapies

This literature research paper focuses on how mutation-specific derived therapies compare to standard treatments in improving clinical outcomes and overcoming therapy resistance in pediatric and adult gliomas. The clinical trials summarized above clearly demonstrate the importance of using mutation-based therapies to control tumor growth while minimizing systemic toxicity. The particular approach has been catalyzed due to the heterogeneity and diversity of genetic mutations such as mIDH1 and BRAF.

4.2 Pediatric Glioma Therapies

For pediatric patients who suffer from gliomas, mutation-based therapies have provided encouraging results, especially for addressing treatment resistance that is most commonly yielded from standard chemotherapy. In the case of BRAF mutations, the BRAF V600E variant is a common cause of tumorigenesis in pediatric low-grade gliomas. However, the drug Dabrafenib, a BRAF inhibitor, combined with either trametinib to suppress the MAPK pathways or the type II RAF inhibitor Tovorafenib to shut down uncontrollable cell growth, has achieved not only tumor shrinkage but also prolonged progression-free survival, outperforming the standard chemotherapy, which typically displays a lower PFS and larger grade adverse events. Most of the low-grade AEs encompass alopecia, fatigue, and mild skin toxicities. Moreover, targeting multiple parts of the MAPK pathway by using a combination therapy further addresses the chances of the body developing resistance to the drugs. Tumors often do this by pathway reactivation or compensatory signaling. However, by using two different inhibitors of both RAF and MAPK, this reduces the likelihood of tumor progression.

Similarly, Everolimus, an mTOR inhibitor, employed the same mechanism for neurofibromatosis type 1-associated gliomas. However, this specific clinical trial focused on

patients whose gliomas had progressed or relapsed even after standard treatment. Therefore, further highlights why mutation-based therapies are vital for prolonging survival. Everolimus effectively disrupted the cell growth and survival by inhibiting the said pathway, ultimately allowing the patients to achieve stable disease and partial response. This pathway-focused medicine exemplifies how understanding tumor biology can aid in employing or developing therapies that address molecular vulnerabilities that existing radiation or chemotherapy is unable to address.

4.3 Adult Glioma Therapies

For adults, IDH1-mutated LGGs originally pose less risk, but eventually are susceptible to progressing into more aggressive tumors. Ivosidenib, an IDH1 inhibitor, has demonstrated efficacy in stabilizing tumor growth and extending the PFS almost three times to what it was, especially for non-enhancing gliomas, since they harbor fewer mutations and are more susceptible to interventions. Despite the significant reduction in the six-month progression after the use of this drug, highlighting its potential, resistance remains a challenge. This is particularly difficult for patients with enhancing gliomas. The safety profile of Ivosidenib showed great promise through low-grade adverse events. Although this successfully allows its long-term usage, prolonged survival benefits require further research.

The early-phase clinical trial that concerned DNX-2401, which is an oncolytic virus injected straight into the tumors with diffuse intrinsic pontine Glioma, showed great clinical promise. It not only had an acceptable safety profile but also ensured tumor shrinkage or disease stabilization. This directly impacted the median survival rate, allowing it to further suppress any expected outcome. After further analysis, it was discovered that DNX-2401 altered the tumor microenvironment and increased anti-tumor activity through T cell infiltration. These results encourage further research into options that can directly target tumor cells and increase immune responses.

In contrast, the study addressing a tumor cell-lysate vaccine derived from Glioma stem cells offered some important insights concerning unconventional therapies connected to immunity. However, there were also some contradictions. The vaccine was well tolerated and even increased the production of CD8+ T cells that specifically attack tumor cells. None of this translated into significant improvement in progression-free survival or EFS between patients receiving neoadjuvant or adjuvant vaccines.

While all of the reviewed trials show encouraging and promising results, there were also quite a few limitations encountered. Some of these concern limitations such as short-term follow-ups, variability in Information due to varying requirements, and primary outcomes/goals. Most importantly, the lack of head-to-head comparisons with standard therapy, and retrospection for this particular paper. Due to the fact that this is a retrospective literature review research paper, the conclusion can only be derived from already existing literature.

Ideally, future research should prioritize large-scale randomized trials that effectively employ mutation-based therapy and compare the existing standard medication to highlight

superiority. Further investigation should be done concerning the combination therapy, as well as enhancing gliomas and their complexity.

5. Conclusion

This literature review has explored recent advances in mutation-specific therapy concerning adult and pediatric patients suffering from glioblastoma, comparing them to the limited conventional treatment options. Across the clinical trials, it is highlighted that tailored therapy can improve not only survival rates, PFS, and EFS, but also reduce treatment-related toxicities, which provides great strength and hope to cancer patients. Among the reviewed interventions, BRAF and MAPK pathway inhibitors, dabrafenib and trametinib provide some of the most promising results, especially for fighting against treatment resistance, alongside Everolimus and Ivosidenib through their increased rate of PFS and response rate. Despite current treatments remaining limited, the rapid progress of research and medicine addressing mutation-based therapy continues to greatly advance. These advances remain as reminders to be open and hopeful about possible new emerging therapies. Future research should mainly focus on this very concept, delving deeper into the reason behind varying results, creating a larger patient population, and directly comparing it to existing treatments to further assess its effectiveness.

Works Cited

Bouffet, Eric, et al. “Dabrafenib plus Trametinib in Pediatric Glioma with BRAF V600 Mutations.” New England Journal of Medicine, vol. 389, no. 12, 2023, pp. 1108–1120. PubMed. Accessed 15 June 2026. https://pubmed.ncbi.nlm.nih.gov/37733309/.

Cherry, Christopher, et al. “Co-located Cell Types Help Drive Aggressive Brain Tumors.” Johns Hopkins Medicine, 7 Feb. 2025. Accessed 12 June 2026. https://www.hopkinsmedicine.org/news/newsroom/news-releases/2025/02/co-located-cell-types-help-drive-aggressive-brain-tumors.

Hargrave, Darren R., et al. “Efficacy and Safety of Trametinib Monotherapy or in Combination with Dabrafenib in Pediatric BRAF V600–Mutant Low-Grade Glioma.” Journal of Clinical Oncology, vol. 41, no. 3, 2023, pp. 664–674. PubMed. Accessed 18 June 2026. https://pubmed.ncbi.nlm.nih.gov/36375115/.

Kilburn, Lindsay B., et al. “The Type II RAF Inhibitor Tovorafenib in Relapsed/Refractory Pediatric Low-Grade Glioma: The Phase 2 FIREFLY-1 Trial.” Nature Medicine, vol. 30, no. 1, 2024, pp. 207–217. PubMed. Accessed 21 June 2026. https://pubmed.ncbi.nlm.nih.gov/37978284/.

Mellinghoff, Ingo K., et al. “Ivosidenib in Isocitrate Dehydrogenase 1–Mutated Advanced Glioma.” Journal of Clinical Oncology, vol. 38, no. 29, 2020, pp. 3398–3406. PubMed. Accessed 21 June 2026. https://pubmed.ncbi.nlm.nih.gov/32530764/.

Ogino, Hideho, et al. “Randomized Trial of Neoadjuvant Vaccination with Tumor-Cell Lysate Induces T Cell Response in Low-Grade Gliomas.” Journal of Clinical Investigation, vol. 132, no. 3, 2022. PubMed. Accessed 19 June 2026. https://pubmed.ncbi.nlm.nih.gov/34882581/.

Pérez-Larraya, Jaime Gállego, et al. “Oncolytic DNX-2401 Virus for Pediatric Diffuse Intrinsic Pontine Glioma.” New England Journal of Medicine, vol. 386, no. 26, 2022, pp. 2471–2481. PubMed. Accessed 20 June 2026. https://pubmed.ncbi.nlm.nih.gov/35767439/.

Ullrich, Nicole J., et al. “A Phase II Study of Continuous Oral mTOR Inhibitor Everolimus for Recurrent, Radiographic-Progressive Neurofibromatosis Type 1–Associated Pediatric Low-Grade Glioma: A Neurofibromatosis Clinical Trials Consortium Study.” Neuro-Oncology, vol. 22, no. 10, 2020, pp. 1527–1535. PubMed. Accessed 21 June 2026. https://pubmed.ncbi.nlm.nih.gov/32236425/.

Delta Machine Learning for Ceiling Temperature Prediction: Bridging Computation and Experiment in Circular Polymer Design By Situo Xing

Abstract

To be recycled, a polymer needs to have a structure that can easily break up. Ceiling temperature, Tc, is an indicator of this. Ceiling temperature is the temperature at which polymerization and depolymerization are balanced, so it shows how a monomer's structure affects a basic tension in design: the material has to be stable while in use but comes apart easily during depolymerization. Tc is the ratio of two quantities that are not equally easy to estimate: the enthalpy and the entropy. This imbalance is what this review discusses. The argument is that delta machine learning can address it: computation predicts the general trend, while a small amount of experimental data corrects the systematic error. First-principles calculations and neural-network models already provide a useful initial estimate of the enthalpy, which a small correction can improve, while the entropy side lacks a base model. Therefore, we propose that polymerization entropy consists of three parts: molecular fragments, chain conformations, and mixing effects. Their contributions can then be fitted using experimental Tc data. Throughout, a sensitivity analysis shows that the model is more useful for comparing monomers within the same family than for making exact predictions across all monomers.

1. Introduction

Polymers have usually been designed to be durable and easy to process, but very few of them are designed to break down into reusable raw materials after they are thrown away. This is why closed-loop, or circular, chemical recycling matters. Circular here means that a used polymer is broken down in a controlled way and its monomers or other clean raw materials are recovered and used again to make new polymer, instead of the material being burned, sent to a landfill, or downgraded into a lower-value product (Quinn et al.). So recyclability depends partly on molecular design and not only on the recycling process. A well-designed polymer stays stable while it is in use but can still be broken down and recovered afterward. To design for this, we need a molecular property that tells us how easily a given structure can come apart, and the ceiling temperature Tc is one such property. The rest of this review looks at how well Tc can be estimated and how much it can actually guide circular polymer design.

1.1 Ceiling Temperature as a Thermodynamic Entry Point

We can describe this balance using the thermodynamics of the monomer–polymer equilibrium. The free energy of polymerization is $\Delta G_p = \Delta H_p - T\Delta S_p$, and the ceiling temperature Tc is the temperature where polymerization and depolymerization are balanced. When the temperature is above Tc, polymerization is no longer favorable and depolymerization happens instead. But even above Tc, whether the monomer is actually recovered still depends on kinetics, catalysis, and the processing conditions (Dainton and Ivin). Tc is also not a fixed number. It changes with concentration, solvent, and other conditions, so we should treat it as a

condition-dependent value. For example, the apparent Tc of γ-butyrolactone changes by more than a hundred degrees between dilute and bulk conditions (Liu et al.; Hong and Chen). Structure affects Tc as well. Ethylene has a very high ceiling temperature, but α-methylstyrene has a much lower one, because steric crowding and a stabilized radical make depropagation easier (Liu et al.). This is why Tc is useful for circular polymer design. It connects structure to a basic tension in design: the polymer should stay stable while it is in use, but still break down afterward so the monomer can be recovered. But Tc only tells us whether depolymerization is thermodynamically possible. It does not tell us whether depolymerization will be fast, whether it will be selective, or whether it will work in a real recycling process (Liu et al.). So the question we care about is not just whether Tc can be computed. We need to know whether it can be computed accurately enough to actually help a design decision.

1.2 The Predictive Question and the Delta-Learning Opportunity

There are so many possible polymers that it is not realistic to make and test them one by one, so computation is used to narrow down the candidates first. But a computed number is not the same as a reliable design decision (Liu et al.). How good a Tc prediction is depends on two inputs, ΔHp and ΔSp, and on how the uncertainties in their estimates influence Tc. Since Tc comes from the balance between these two, we cannot look at their errors separately. Machine learning relies on a useful fact from computational thermochemistry. A method can give unreliable absolute values but still put related molecules in the right order, because a systematic error usually shifts similar structures by a similar amount (Bartel et al.). This is well known in chemical computations, where the change in energy between reactants and products is typically much more reliable than the energies of the individual components in the reaction. Delta machine learning uses the same idea. It writes the accurate value as a low-cost computed estimate plus a learned correction, $E_{actual} = E_{computed} + \Delta$. The computed values capture the general trends, while the Δ from a small number of high-quality measurements corrects the offset (Ramakrishnan et al.). However, computational estimates of enthalpy and entropy are not at the same stage of development. Enthalpy already has fast and accurate energy models, so the approach works fairly well there. Entropy is the harder one, because there is still no equivalent low-cost model for this quantity.

1.3 Scope, Objective, and Contribution

This review treats Tc as a thermodynamic parameter that can help predict and guide circular polymer design. It does not try to cover every depolymerization mechanism or recycling method. Instead, it focuses on three questions. What can Tc tell designers, and what can it not tell them? How do current methods estimate the two components behind Tc, the enthalpy and the entropy? And can a delta-learning approach combine these estimates into a prediction that designers can actually use? Computing a number is the easy part. What we really want to know is whether that number is accurate enough to guide a design choice. Because entropy is the harder part, we go one step further and propose a base model for it. We split polymerization

entropy into three components and combine them with a linear delta model. We present this as a direction worth testing. We have not built or tested the model ourselves.

2. Structural Basis and the Design Window

2.1 Structural Factors Shaping Tc

There is no single property that decides the ceiling temperature. Tc depends on how the monomer and the polymer together share the energy gain and the entropy cost of polymerization. Many structural features influence Tc, such as ring strain, steric crowding, restricted conformations, substituent electronics, heteroatoms, and the backbone itself. But each of these features changes both ΔHp and ΔSp at the same time, and the effect depends on the monomer family, so there is no rule that works for every case (Liu et al.; Olsén et al.). Ring strain is the most obvious factor for cyclic monomers, but it is not enough by itself. γ-Butyrolactone and δ-valerolactone have about the same ring strain, roughly 8 kcal/mol, but the first one does not polymerize under normal conditions while the second one does. Quantum-chemical calculations trace this difference to two things: the ester group only has to distort a little in the five-membered ring, and the polyester chain that forms is conformationally stable (Houk et al.). The fused-ring cyclooctenes studied by Zhou, Sathe, and Wang make this even clearer. In this family the polymerization enthalpy changes very little, but the ceiling temperatures at 1 M range from 330 to 680 °C. This large range is caused mainly by entropy. In fact, gem-disubstituents can change Tc by about 300 °C even though they are not on the reacting ring (Zhou et al.). Table 1 lists representative values.

Table 1. Representative structure–thermodynamics relationships relevant to circular polymer design. Values are as reported or compiled in the cited sources.

Monomer family	**Structural variable**	**Thermodynamic evidence (as reported)**	**Key design insight**
Unsubstituted lactones (4–7-membered)	Ring size	ΔHp ≈ −82 to −5 kJ/mol; ΔSp ≈ −29 to −54 J/(mol·K) (Olsén et al.)	Ring size systematically shifts ΔHp and ΔSp; the 5-membered ring is an outlier
γ-butyrolactone vs δ-valerolactone	5- vs 6-membered ring	Comparable ring strain ≈ 8 kcal/mol; γ-BL Tc ≈ −136 °C at 1 M (Hong and Chen; Houk et al.)	Ring strain alone is insufficient; conformational stability and ester distortion matter
Fused-ring cyclooctenes	Fused ring size; remote gem-disubstitution	ΔHp = −2.1 to −3.3 kcal/mol; ΔSp = −2.7 to −5.0 cal/(mol·K); Tc = 330–680 °C at 1 M (Zhou et al.)	ΔSp, not ΔHp, drives the large Tc variation; remote substituents shift Tc by ≈ 300 °C
Thiolactones	β-substituent; sulfur heteroatom	For a key monomer, ΔHp = −19.3 kJ/mol; ΔSp = −58.4 J/(mol·K) (Stellmach et al., “Modulating”)	Substituent and backbone heteroatom effects are tunable; conversion can be temperature-sensitive

Gem-disubstituted δ-valerolactones	gem-α,α-dialkyl substitution	Complete chemical recyclability; some variants report Tm up to 140 °C (Li et al.)	Thorpe–Ingold-type design can enhance recyclability while keeping useful performance

Together, these examples show something specific but limited. Structure changes Tc through combined enthalpy and entropy effects, and a structural change is most reliable as a guide when we stay inside one monomer family (Liu et al.; Olsén et al.; Stellmach et al., "Modulating"). Olsén and co-workers show that ΔHp and ΔSp change in a regular way as the lactone ring size changes, with the five-membered ring as the exception (Olsén et al.). Thiolactones show the same kind of behavior through substituents and a sulfur heteroatom, and DFT conformational analysis supports this (Stellmach et al., "Modulating").

2.2 Screening and the Stability–Recoverability Window

Because these effects are clearest within related scaffolds, this kind of analysis is most useful before synthesis. Reasoning from structure to Tc will not give a designer a final answer, but it can narrow the search by pointing to scaffolds where polymerizability and recoverability are more likely to go together. The safe way to use it is to rank related candidates within a family and remove the ones that clearly will not work, not to predict exact Tc values across all of chemistry. A substitution that lowers Tc in one family may do nothing, or even the opposite, in another family (Liu et al.; Olsén et al.; Zhou et al.).

It also helps to think about a workable range instead of a single target value, because neither extreme is good. If Tc is very high, the polymer is stable, but the monomer is almost impossible to recover. If Tc is very low, depolymerization is easy, but the polymer becomes hard to make and use. γ-Butyrolactone shows the problem with the low end. Its Tc is about −136 °C at 1 M, which makes closed-loop recycling look attractive on paper, but it forces the polymerization to run at low temperature and high concentration, which is demanding (Hong and Chen). Gem-disubstituted valerolactones and hybrid monomers show the other option, where tuning Tc toward the middle lets recyclability and good performance exist together (Li et al.; Shi et al.). So the real goal is to find a workable range, wherever the structure and the conditions place it, rather than pushing Tc as high or as low as possible (Liu et al.).

3. The Delta-Learning Framework and Enthalpy Estimation

3.1 Machine Learning and Delta Models in Chemistry

Most machine-learning models in chemistry are built in two steps, and energy prediction is the clearest example. In the first step, a high-cost but accurate method, usually density functional theory (DFT), is used to create a large reference dataset. In the second step, a cheap model is trained to reproduce those results at a much lower cost. These datasets have grown quickly. ANI-1 contains more than 20 million off-equilibrium DFT conformations for 57,462 small organic molecules, and it was built to train transferable neural-network potentials (Smith et

al., “ANI-1”). Later, active learning reached a similar accuracy with much less repeated data by choosing the most informative configurations as the training went on (Smith et al., “Less Is More”). OMol25 is even larger, with more than 100 million DFT single-point calculations, 83 elements, and systems of up to 350 atoms, which took about six billion CPU core-hours to produce (Meta FAIR).

Underneath, the model is learning a function that maps features to a target, written as $f(\text{Feature}_1, \text{Feature}_2, \ldots) \rightarrow \text{Target}$. The features describe the structure or the local environment of the atoms, and the target is the property we want. For energy specifically, the model usually writes the total energy as a sum of atomic energies, and each atomic energy is predicted from features that describe the local environment of that atom. The simplest version is a linear model, $C_0 + C_1 \cdot \text{Feature}_1 + C_2 \cdot \text{Feature}_2 + \ldots = \text{Target}$. Training the model means adjusting the coefficients so the predictions match the reference values as closely as possible. The neural-network potentials used for ANI-1 and OMol25 replace this linear fit with a flexible nonlinear function that has many parameters, but the basic idea is the same. It is still a function from features to a target, with parameters fitted to data. The second step builds directly on this.

3.2 Estimating Enthalpy: Mature Models and Residual Delta-Correction

Polymerization enthalpy comes from the bonds that are broken and formed when a monomer becomes part of a chain, so ΔH_p fits well with reaction-energy methods. Between the two inputs to Tc, enthalpy is the easier one to get. This is partly because the energy is mostly set by a small number of low-energy structures, so an optimized geometry is often enough and we do not need the full set of conformations. This does not mean enthalpy is just ring strain. It depends on what the monomer loses and what the polymer gains, and substituents, heteroatoms, and ring fusion all change it in ways that matter inside a family but cannot be reduced to one number (Houk et al.). When we use these methods to compare candidates, they are good at telling us which of two related monomers is more favorable.

These methods are now accurate enough to be practical. One method that handled the conformational search and finite-size effects carefully reached a root-mean-square error of about 7 kJ/mol for ΔH_p across 42 different ring-opening systems, which is about three times better than the usual approach (Tran et al.). Machine-learning models trained on a mix of experimental and simulation data are now getting close to this accuracy at a much lower cost (Toland et al.). General-purpose neural-network potentials such as AIMNet2, which come from the same line as the ANI models and datasets like OMol25, also speed up the geometry optimization and conformer search that the energy estimates need (Anstine et al.). Together, these give a reasonable trend-level estimate for enthalpy, which is exactly the input that a delta-correction is meant to improve.

Still, these methods are not exact, and it helps to be clear about where they fall short. The energy models are often fitted on idealized or gas-phase references, they treat solvation only roughly, and they have to go from small fragments to real polymers, condensed phases, and chain-length effects. Each of these creates a systematic gap between the computed ΔH_p and the

value that actually applies under real conditions. This gap is what the delta-correction is for. We write ΔHp,actual = ΔHp,ML + Δ_H, give Δ_H a few parameters, and fit them using reference or experimental data. The base model keeps the trend correct, and the correction absorbs the systematic error. So enthalpy already has what the framework needs: a reasonable trend-level estimate and a simple correction with few parameters. Entropy does not have this yet. In the next section we discuss the challenges of building a similar approach for the entropy of polymerization.

4. Entropy Estimation: The Central Challenge

4.1 Why Entropy Is the Harder Term

Polymerization almost always lowers entropy. Many monomers that could move freely on their own become part of one chain, so they lose most of their translational freedom. This is usually why polymerization becomes less favorable as the temperature rises, even when ΔHp is negative (Duda and Kowalski; Hodge). But this is only part of the story. The net ΔSp is not just the translational loss. It is the translational loss together with the change in conformational freedom when the chain forms, and this second part can go in either direction. Some strainless macrocycles even polymerize because of entropy, not against it. When a constrained ring opens into a flexible chain, it can gain more conformational freedom than it loses in translation, so ΔHp is near zero and ΔSp is positive (Hodge). This means entropy is specific to each system. A model for it has to follow chain flexibility and ring shape, not just count molecules.

This is the deeper reason entropy is harder than enthalpy. Energy is mostly set by the lowest-energy structures, so a few optimized geometries can represent the whole system. Entropy is different. It depends on the number and the weight of all the states the system can reach, so it is sensitive to how completely and how fairly we sample those states, and we cannot get it from a single structure (Hodge; Peter et al.). A change that moves ΔHp by a clear, local amount can move ΔSp through a small, spread-out shift in many conformers that no single structure shows. For enthalpy we have DFT, which is cheap enough, physically based, and correct about trends. The rest of this section asks what could play the same role for entropy.

4.2 A Three-Component Decomposition of Entropy

We propose to give entropy a base model by splitting the polymerization entropy into three parts that add up:

$$\Delta Sp \approx \Delta S_{fragment/group} + \Delta S_{conformational/RIS} + \Delta S_{mixing/standard\text{-}state}$$

Each part has a clear physical meaning and its own way of being estimated. The first part is the fragment or group term. This is the contribution of the local chemical groups to the entropy of the monomer and the polymer. The entropy from bonds, rotating groups, and substituents is roughly the same between molecules that share the same local environments, so we can estimate this part by adding up tabulated group contributions. This is the same method used to build

standard entropies of formation from group values. This part covers the share of ΔSp that follows the chemical composition rather than the overall shape of the chain.

The second part is the conformational term. This is the change in the chain's conformational freedom during polymerization, and it is the part that is hardest to describe in a local way. The natural tool for it is rotational-isomeric-state (RIS) theory. RIS is a method from classical statistical mechanics, developed by Flory and later extended for polymers, that treats the chain as a set of discrete torsional states whose populations and short-range correlations decide the conformational statistics (Flory). RIS fits here because polymerization does more than join molecules together. It changes the size, the connectivity, and the flexibility of the available conformational space, and a torsional-state description is built to track exactly this change.

The third part is the mixing and standard-state term. This part has nothing to do with structure and everything to do with how we count: the entropy of mixing and the conventions that define the reference state. These are not small details. As Section 5 explains, the apparent Tc of one monomer can change by more than a hundred degrees depending on the concentration and the standard state used to define equilibrium. So if we want a predicted entropy to match a measured one, we need a term that carries this dependence directly.

This split does for entropy what DFT does for enthalpy. It provides a base model we can interpret, built from quantities that can each be computed or looked up. Instead of treating ΔSp as one number that we pull indirectly from experiment, we separate it into three parts with different origins and different methods. This way, the parts we understand well, the group contributions and the standard-state corrections, are not held back by the part we understand least, the conformational term. The largest source of uncertainty is now visible instead of hidden, and these three parts are exactly what a delta model can take and calibrate.

4.3 Estimation Methods per Component and a Proposed Linear Delta Model

Each part has an established method behind it, and the differences between them are the point. The fragment or group term fits Benson-type group additivity, an established method that builds thermodynamic properties, including entropy and heat capacity and not just enthalpy, by adding up transferable group contributions fitted to a large amount of data (Benson and Buss). Its strengths are speed and coverage. The set of chemical fragments that actually appear in polymerizable monomers is relatively small, so a group-contribution table can cover most of the cases a screening model needs to look at. Its weakness is that it captures the local, composition-based entropy but misses the long-range conformational statistics of the chain. That long-range part is the job of the second term.

The conformational term comes from RIS, which obtains the conformational partition functions and chain dimensions from the statistics of the torsional states (Flory). RIS gives a low-cost estimate, which is what a screening model needs. When the torsional behavior of a system is too specific for tabulated parameters, molecular dynamics (MD) can in principle fill the gap, but MD is computationally expensive and does not scale well to scanning many candidate structures. Getting a reliable entropy out of MD depends on generating trajectories that

actually sample the relevant conformations, and long chains relax slowly while condensed-phase packing adds correlations that a gas-phase picture misses, so this sampling is hard (Peter et al.; Gartner and Jayaraman). For those reasons we do not build the method on MD. We keep the conformational term at the low-cost RIS level and let the delta calibration described below absorb the part it gets wrong. This is also why the conformational term, more than the other two, is the one that most needs to be calibrated against experiment.

The mixing and standard-state term is handled with the standard solution-thermodynamics treatment of concentration and reference state, which is well established even though the literature does not always apply it consistently (Duda and Kowalski). Since these three estimates are each only partly reliable, a simple starting point is to combine them in a linear model:

$$\Delta Sp \approx c_0 + c_1 \cdot \Delta S\_fragment/group + c_2 \cdot \Delta S\_conformational/RIS + c_3 \cdot \Delta S\_mixing/standard\text{-}state$$

The constant term c_0 absorbs a fixed offset, and the coefficients c_1, c_2, and c_3 would be learned by fitting to experimental Tc, so that the fit can weight each part by how reliable it is for a given family and soak up its systematic error. This linear model is a simple delta model. The three components give a base prediction with the correct physical trends, and the fitted weights would pull that prediction toward the measured behavior. We are not going to develop this model in this review. We only want to point to a path that reaches ΔSp from methods that are already well established, so that entropy can enter a delta-corrected route to Tc the same way enthalpy does. In short, the three components give a cheap and interpretable base estimate for entropy, a constant and three weights would calibrate it, and together with the enthalpy base model and its correction from Section 3 they would form a complete delta-corrected route to Tc. We present this as a direction for future work that should be tested against data, not as a finished method.

5. From H and S to Tc: Sensitivity, Reliability, and Experiment

5.1 Combining Estimates for H and S

Once we have estimates for both terms, we can get Tc only after we make the equilibrium assumptions explicit. In the simplified standard-state form, setting $\Delta Gp = 0$ gives $Tc \approx \Delta Hp/\Delta Sp$ (Duda and Kowalski). So far we have described separate delta models for enthalpy and for entropy. But we can also build a single delta model that takes the computed components of both ΔHp and ΔSp as inputs and predicts Tc directly, with its parameters obtained from a fit to experimental Tc. Fitting against Tc has a few practical advantages: Tc is easy to measure, so it can constrain the enthalpy and entropy parameters at the same time, it gives the enthalpy and entropy errors a chance to partly cancel inside the fit, and it calibrates the model against the exact quantity a designer cares about.

We can use a simple sensitivity analysis to see how the errors in ΔHp and ΔSp combine, and whether they cancel or multiply. Propagating the uncertainty through $Tc = \Delta Hp/\Delta Sp$ gives

roughly $\sigma(T_c)^2 \approx [1/\Delta S_p]^2\sigma(\Delta H_p)^2 + [\Delta H_p/\Delta S_p^2]^2\sigma(\Delta S_p)^2$. ΔS_p sits in the denominator of both terms, so when $|\Delta S_p|$ is small, a given error in entropy turns into a large error in Tc. This amplification is not a limitation of any particular model. It is built into the relation between entropy and Tc, and it gets worse exactly where $|\Delta S_p|$ is small, which is the low-Tc, near-thermoneutral region that circular design cares about most. Liu and co-workers show the same picture from the other direction: across much of the useful design space, the uncertainty in both ΔH_p and ΔS_p has to become impractically small before Tc can separate close candidates, and the precision we need is not one fixed number but depends on where the candidate sits in ΔH_p–ΔS_p space and on how strict the design question is (Liu et al.). A concrete example shows how thin the margin can be. In one van 't Hoff analysis, a shift in ΔS_p° of only 5.7 J/(mol·K) moves the calculated Tc by 37 °C, so a small inconsistency in how the reference state is assigned already changes the answer by tens of degrees (Stellmach et al., “Improving”).

The difference between the two terms is the main lesson for design. ΔS_p is usually smaller in magnitude and less accurately known than ΔH_p, and it is often obtained indirectly, by fitting equilibrium monomer concentrations against temperature in a van 't Hoff analysis rather than from first principles (Schwartz et al.; Stellmach et al., “Improving”). Because ΔS_p enters Tc through the denominator, its uncertainty has a larger effect on the result. This is why a good entropy base model, the one proposed in Section 4, is the central problem and not a side issue. A recent machine-learning study shows where the field stands: trained on the largest experimental dataset of polymerization thermodynamics so far, it predicted ΔH_p and ΔS_p with reasonable errors of about 1.27 kcal/mol and 3.24 cal/(mol·K), yet these combined into a Tc error of about 136 °C (Kozarekar et al.). Getting each part acceptably right does not guarantee a useful ratio, which is the amplification described above showing up in real numbers. A delta model calibrated directly on Tc is meant to attack exactly this, by tuning the corrections against the combined quantity that design decisions actually depend on.

5.2 Absolute Values, Relative Trends, and Design Usefulness

The literature often treats “predicting Tc” as one task, but it is really two different tasks. One is absolute prediction, meaning an accurate Tc for any monomer we are given, and current methods usually cannot do this. The other is trend prediction, meaning ranking related candidates within a family, which is narrower and is usually what early design needs. A method can do the second well and still fail the first, because a systematic error that shifts a whole family by the same amount keeps the ranking correct even when the individual numbers are wrong (Bartel et al.; Tang et al.). The delta framework reflects this difference directly. The base model is responsible for the trend, and the delta parameters absorb the error in the absolute scale.

This changes what we should mean by “good enough.” Instead of judging a method only by how close it is to the absolute value, the better question is whether the remaining uncertainty is small enough for the decision in front of us. That decision might be removing candidates that clearly will not work, finding which structural change moves Tc in the right direction, ranking a family, or checking whether a candidate falls inside or outside a target range. These tasks need

very different levels of precision. Judging a method by how useful it is for design, rather than by raw accuracy, gives a more honest picture of where computation already helps in circular polymer design (Liu et al.).

5.3 What Tc Does Not Cover

Computation cannot close the loop on its own, and neither can Tc. Real recyclability also depends on kinetics, catalysis, processing, and phase behavior, none of which Tc describes, and the apparent Tc itself shifts with concentration, solvent, and thermal history (Lohmann et al.; Stellmach et al., "Improving"). So a favorable Tc is a necessary starting point, not a guarantee that a monomer can be recovered in practice. This review addresses only the Tc side of the problem. Even there, the experiment is not just a final check in the delta framework. It is the training signal that sets the corrections, and it is the test of whether a prediction means anything for design, which is why reporting the concentration, the solvent, and the thermal protocol in a standard way is necessary and not optional (Stellmach et al., "Improving"; Lin et al.).

6. Discussion and Outlook

The idea this review builds toward is simple to state. We use low-cost computation to estimate the factors that go into ΔH_p and ΔS_p, turn those estimates into features, and let a delta model calibrated on experimental Tc turn them into a prediction. In practice this would run in a few steps. We first do a rough screen with ML-assisted enthalpy, which is already fast and reasonably accurate. We then build the three-component entropy estimate from group contributions, RIS, and the standard-state term. We obtain the delta parameters by fitting them to experimental Tc, and we use the sensitivity analysis to decide how precise each decision actually has to be. Finally, we compare the result against the experiment to place each candidate inside a workable range, and we repeat as more data come in. Even this is limited by the state of the data, because registration barriers and inconsistent reference-state conventions still make thermodynamic results hard to combine (Lin et al.). So the workflow is a reasonable conclusion from the current literature, not something anyone has yet run from start to finish.

Whether this works depends on the enthalpy–entropy asymmetry that runs through the whole review. Enthalpy is becoming easier to reach and to interpret, because there are now mature first-principles benchmarks, neural-network potentials, and a delta-correction that cleans up the remaining bias. Entropy is still harder, less standardized, and more dependent on sampling and reference-state choices, and the three-component decomposition is our attempt to give it the same kind of base model that already works for enthalpy. As a framework this has real advantages: it leans on computed trends that stay reliable even when the absolute values are not, it needs relatively little experimental data, each term keeps a clear physical meaning, and it can be improved family by family as more data appear. The limits are just as real. A linear combination may be too simple when the components are strongly coupled, we do not know how well the fit transfers between very different families, and that transfer is held back by the same error-cancellation that makes trend prediction work in the first place. Three components may

also miss something that matters in a specific chemistry, and experimental Tc values are sometimes reported under conditions inconsistent enough that a naive fit would learn the inconsistency instead of correcting it (Tang et al.). For all these reasons this is a direction to test, not a settled method.

For practice, the conclusion is straightforward. The most reliable use of Tc computation right now is not to chase exact universal values, but to narrow the design space in a sensible way, by comparing structures, by reading the trade-off between stability and recoverability, and by going back and forth with experiment. Tc is a genuinely useful thermodynamic anchor for circular polymer design, but how useful it is depends on how well we can estimate and interpret its enthalpy and entropy terms, and those two are not equally developed. Delta machine learning is a consistent way to bridge that gap, pairing a mature enthalpy base model and a small correction with the proposed three-component entropy model and its linear delta, and fixing both against experimental Tc. We should still state the limits of this review plainly: it addresses only Tc, the three-component scheme has not been validated, and the proposed fitting and its transferability still need to be tested. So the most promising near-term direction is prediction that is anchored to experiment, aware of its own calibration, and built for design, rather than another purely numerical demonstration.

Works Cited

Anstine, D. M., et al. “AIMNet2: A Neural Network Potential to Meet Your Neutral, Charged, Organic, and Elemental-Organic Needs.” Chemical Science, vol. 16, 2025, pp. 10228–44.

Bartel, C. J., et al. “A Critical Examination of Compound Stability Predictions from Machine-Learned Formation Energies.” npj Computational Materials, vol. 6, 2020, article 97.

Benson, S. W., and J. H. Buss. “Additivity Rules for the Estimation of Molecular Properties: Thermodynamic Properties.” The Journal of Chemical Physics, vol. 29, no. 3, 1958, pp. 546–72.

Dainton, F. S., and K. J. Ivin. “Reversibility of the Propagation Reaction in Polymerization Processes and Its Manifestation in the Phenomenon of a ‘Ceiling Temperature.’” Nature, vol. 162, 1948, pp. 705–07.

Duda, A., and A. Kowalski. “Thermodynamics and Kinetics of Ring-Opening Polymerization.” Handbook of Ring-Opening Polymerization, edited by P. Dubois, O. Coulembier, and J.-M. Raquez, Wiley-VCH, 2009, pp. 1–51.

Flory, P. J. Statistical Mechanics of Chain Molecules. Interscience, 1969.

Gartner, T. E., and A. Jayaraman. “Modeling and Simulations of Polymers: A Roadmap.” Macromolecules, vol. 52, 2019, pp. 755–86.

Hodge, P. “Entropically Driven Ring-Opening Polymerization of Strainless Organic Macrocycles.” Chemical Reviews, vol. 114, 2014, pp. 2278–312.

Hong, M., and E. Y.-X. Chen. “Completely Recyclable Biopolymers with Linear and Cyclic Topologies via Ring-Opening Polymerization of γ-Butyrolactone.” Nature Chemistry, vol. 8, 2016, pp. 42–49.

Houk, K. N., et al. “Why δ-Valerolactone Polymerizes and γ-Butyrolactone Does Not.” The Journal of Organic Chemistry, vol. 73, no. 7, 2008, pp. 2674–78.

Kozarekar, S. S., et al. “Predicting Ceiling Temperature for Recyclable Polymer Design Using Machine Learning.” Cell Reports Sustainability, vol. 3, no. 4, 2026, article 100652.

Li, X.-L., et al. “A Circular Polyester Platform Based on Simple Gem-Disubstituted Valerolactones.” Nature Chemistry, vol. 15, no. 2, 2023, pp. 278–85.

Lin, T.-S., et al. “PolyDAT: A Generic Data Schema for Polymer Characterization.” Journal of Chemical Information and Modeling, vol. 61, no. 3, 2021, pp. 1150–63.

Liu, X., et al. “Understanding Ceiling Temperature as a Predictive Design Parameter for Circular Polymers.” Cell Reports Physical Science, vol. 5, no. 4, 2024, article 101910.

Lohmann, V., et al. “The Thermodynamics and Kinetics of Depolymerization: What Makes Vinyl Monomer Regeneration Feasible?” Chemical Science, vol. 15, 2024, pp. 832–53.

Meta FAIR. “The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models.” arXiv, 2025, arXiv:2505.08762. Preprint.

Olsén, P., et al. “Thermodynamic Presynthetic Considerations for Ring-Opening Polymerization.” Biomacromolecules, vol. 17, no. 3, 2016, pp. 699–709.

Peter, C., et al. “Estimating Entropies from Molecular Dynamics Simulations.” The Journal of Chemical Physics, vol. 120, 2004, pp. 2652–61.

Quinn, E. C., et al. “Cracking the Code of Multi-Layer Films to Promote Circularity in Single-Use Plastic Packaging.” Nature Communications, vol. 17, 2026, article 1489.

Ramakrishnan, R., et al. “Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach.” Journal of Chemical Theory and Computation, vol. 11, 2015, pp. 2087–96.

Schwartz, J. M., et al. “Determination of Ceiling Temperature and Thermodynamic Properties of Low Ceiling Temperature Polyaldehydes.” Journal of Polymer Science Part A: Polymer Chemistry, vol. 56, 2018, pp. 221–28.

Shi, C., et al. “Hybrid Monomer Design for Unifying Conflicting Polymerizability, Recyclability, and Performance Properties.” Chem, vol. 7, no. 3, 2021, pp. 670–85.

Smith, J. S., et al. “ANI-1, a Data Set of 20 Million Calculated Off-Equilibrium Conformations for Organic Molecules.” Scientific Data, vol. 4, 2017, article 170193.

Smith, J. S., et al. “Less Is More: Sampling Chemical Space with Active Learning.” The Journal of Chemical Physics, vol. 148, no. 24, 2018, article 241733.

Stellmach, K. A., et al. “Improving the Accuracy of Ceiling Temperature Measurements: Best Practices and Common Pitfalls.” Macromolecules, vol. 58, no. 8, 2025, pp. 3729–41.

Stellmach, K. A., et al. “Modulating Polymerization Thermodynamics of Thiolactones through Substituent and Heteroatom Incorporation.” ACS Macro Letters, vol. 11, no. 7, 2022, pp. 895–901.

Tang, H., et al. “Assessing Uncertainty in Machine Learning for Polymer Property Prediction: A Benchmark Study.” Journal of Chemical Information and Modeling, vol. 65, 2025, pp. 6585–98.

Toland, A., et al. “Accelerated Scheme to Predict Ring-Opening Polymerization Enthalpy: Simulation–Experimental Data Fusion and Multitask Machine Learning.” The Journal of Physical Chemistry A, vol. 127, no. 50, 2023, pp. 10709–16.

Tran, H., et al. “Toward Recyclable Polymers: Ring-Opening Polymerization Enthalpy from First-Principles.” The Journal of Physical Chemistry Letters, vol. 13, 2022, pp. 4778–85.

Zhou, J., et al. “Understanding the Structure–Polymerization Thermodynamics Relationships of Fused-Ring Cyclooctenes for Developing Chemically Recyclable Polymers.” Journal of the American Chemical Society, vol. 144, no. 2, 2022, pp. 928–34

Developing Sustainable Growth Policies: A 2035 Policy Transfer Model for Loudoun County and Kaiserslautern By Sarina Virmani

Abstract

As cities continue to grow around the world, communities are faced with an increasingly relevant challenge: maintaining sustainable growth. This project explores how sustainable growth can be optimized in two distinct areas, Loudoun County, U.S.A. and Kaiserslautern, Germany. Portland, U.S.A. and Freiburg, Germany were used as benchmark sustainable cities. Each place's current sustainable situation was scored across five categories: transportation, land use, energy, water, and governance to determine a cumulative Sustainable Growth Score for each city. Models were created to determine the expected Sustainable Growth Score in 2035 under three additional scenarios: Business as Usual, Moderate Reform, and Policy Transfer. The Policy Transfer model represented a world in which Loudoun and Kaiserslautern adopted aspects of policies from other cities and adapted them to fit local conditions to improve their respective sustainability scores. Results found that Loudoun should focus on improving its transportation and energy sustainability, specifically by implementing low-car districts and mandating large-load users to submit clean-energy plans as well as expand clean energy transmission. Kaiserslautern performed better than Loudoun but could still focus on improving its energy and water sustainability categories by closing bike and public transit network gaps and expanding de-sealing as well as rainwater retention.

1. Introduction

Urbanization is accelerating around the world. As of 2024, over half of the world's population live in urban areas. By 2050, nearly 7 out of 10 people are expected to live in urban areas (United Nations). As people migrate to cities, technology advances, and infrastructure expands, sustainable growth has become one of the most prominent planning challenges. Communities are expected to balance economic development, affordable housing, accessible transportation, energy resources, and climate resilience. Without proper governance and planning, maintaining this balance can be challenging.

However, growth does not affect every place in the same way. A car-dependent county in the United States facing rapid data center and infrastructure expansion, for example, faces different sustainability challenges than a compact European city with comprehensive land-use planning. Because of these inherent differences, policies cannot simply be copy and pasted from one place to another. Rather, they have to be translated to fit the needs and infrastructure of each community. This study explores this tension by examining how policies can be adapted and transferred from one place to another to facilitate sustainable growth. Specifically, it compares two areas with distinct growth pressures: Loudoun County, U.S.A. and Kaiserslautern, Germany.

Loudoun County is a rapidly growing suburb in Virginia, United States. One of the biggest challenges to sustainable growth is the proliferation of data centers and the large amount of resources they require. Namely, Loudoun hosts over 200 data centers that take up over 45 million square feet of land. Other than land, these data centers require massive amounts of water

(millions of gallons each day) and energy, which is beginning to strain the community (Faguy). Loudoun has responded through policies such as transportation planning, zoning standards, limited data center regulation, energy planning, and natural-resource efforts. However, many of these policies have limitations that allow data centers to expand uncontrollably with limited regard for sustainability. Loudoun is useful to study as it is a place under high growth pressure that could benefit from policy-transfer strategies. On the other hand, Kaiserslautern, Germany, faces different sustainability challenges. It already has a strong foundation for planning in several categories. For example, it already has transportation planning that emphasizes walking, biking, and public transportation expansion. Additionally, it has land-use planning that supports compact growth and brownfield reuse. However, Kaiserslautern faces implementation challenges that originate, at least in part, from policy language that could be strengthened. For instance, they struggle with connecting sustainability goals to funding and action.

To make the comparison more useful, the study also uses Freiburg, Germany and Portland, Oregon as benchmark sustainable cities. Freiburg is often regarded as the most sustainable city in Germany. Freiburg has low reliance on private cars, an extensive cycling network, compact district planning, homes that maximize energy-efficiency, and green-blue infrastructure (McDarris). Portland is a helpful U.S. benchmark for sustainability as it prioritizes renewable energy generation, waste reduction, public accountability, strict land-use policies, and public transportation (Sense). While these cities are not perfect at every aspect of sustainability, they offer a means for comparison for what an ideal sustainable city might look like. Additionally, these cities do not face the same exact challenges as Loudoun or Kaiserslautern. Instead, the policies from Portland and Freiburg can help define what feasible improvements could look like when adapted to the conditions of Kaiserslautern or Loudoun County.

This comparative approach draws roots from existing research. Studies find that sustainable urban development can not be accurately measured through a single factor, but rather through frameworks that combine various categories. This supports the use of the multi-category growth score in this study (Michalina et al.). Another study that examined SDG 11 indicators in Germany found that global sustainability goals have to be adapted to local contexts to be successful in implementation. Therefore, the policies employed to improve sustainability must be adapted to fit each community's unique conditions (Koch and Krellenberg).

Thus, the research question of this study is: How can Loudoun County and Kaiserslautern improve sustainable growth outcomes by 2035 by adapting policies from U.S. and German benchmark cities? To address this question, the study develops a scenario-based policy implementation model. Each place is scored from 1 to 5 across five sustainability categories: transportation, land use, energy, water, and governance. These categories are then weighted to determine a Sustainable Growth Score for each city. Transportation and land use are weighted most heavily. They shape demand for infrastructure, affecting urban development in the long-term. The model then explores four scenarios: current conditions, Business as Usual 2035, Moderate Reform 2035, and Policy Transfer 2035. Current conditions represent the baseline world. Business as Usual 2035 represents a world in which Loudoun and Kaiserslautern's

policies remain the same. Moderate Reform 2035 represents a world in which Loudoun and Kaiserslautern adopt moderate reforms and strengthen the policies that already exist. Policy Transfer 2035 represents a world in which Loudoun or Kaiserslautern adopts policies from the other cities and adapts these policies to improve their sustainability score.

2. Policy Context

This section summarizes the policies that exist in each city or region across the categories in the sustainable growth model. Governance is not included in this section, as the policies represent governance.

2.1 Transportation Policies

The purpose of this category is to see whether the city or region has taken meaningful steps to lower car dependence and support low-emission mobility (walking, biking, public transit).

Table 1. Transportation Policies for Each City/Region

Place	Main Policies	What it does	Implementation Challenges
Loudoun County	2019 Countywide Transportation Plan	Encourages public transportation options, mixed-use development, ridesharing, EV charging, and lower vehicle miles traveled	Implementation depends on funding, development proposals, and VDOT coordination (Loudoun Government)
Kaiserslautern	Mobilitätsplan Klima 2030	Lowers car dependence and encourages walking, biking, and public transit; identifies bike-network gaps and recommends stronger bus/rail connections	Implementation depends on funding, street redesign, public acceptance, and coordination with transit providers (Stadt Kaiserslautern)
Freiburg	VEP 2020 and Cycling Concept 2020	Builds long-term low-car mobility through cycling, rail, short distances, separated bike routes, and bike parking	Difficult to copy in spread-out suburban regions (ICLEI) (Kotas)
Portland	2035 Transportation System Plan	Prioritizes walking, biking, and transit; uses transportation demand management and electric vehicle infrastructure	Implementation depends on funding and agency coordination (Portland Government)

2.2 Land Use Policies

The purpose of this category is to see whether the city or region has taken meaningful steps to guarantee sustainable zoning and green space.

Table 2. Land Use Policies for Each City/Region

Place	Main Policies	What it does	Implementation Challenges
Loudoun County	Zoning Ordinance + Phase 1 Data Center Regulation	Adds green space, tree canopy, landscaping, facades, setback, and generator standards; Phase 1 regulation moved data centers from automatic (by-right) approval to Special Exemption review in industrial districts	Emphasized location and site design over setting sustainability standards (Loudoun Government)
Kaiserslautern	Flächennutzungsplan 2025	Prioritizes inner development, brownfield reuse, compact growth, and development near transit corridors	Still allows new settlement expansion, depends on feasibility of redevelopment plans (Stadt Kaiserslautern)
Freiburg	Land Use Plan 2020/2040 + Dietenbach/Vauban model	Supports compact mixed-use neighborhoods, affordable housing, landscape planning, and sustainable land allocation.	Difficult to copy in less compact places (Green City Freiburg) (Urban Agency)
Portland	Zoning Code Title 33 + Residential Infill Project	Guarantees tree protections, inclusive housing, eco-friendly roof requirements, and middle housing reforms	Protections vary by context and implementation depends on enforcement and local conditions (Portland Government)

2.3 Energy Policies

The purpose of this category is to determine whether the city or region has taken meaningful steps to transition to low-emission forms of energy.

Table 3. Energy Policies for Each City/Region

Place	Main Policies	What it does	Implementation Challenges
Loudoun County	2023 Energy Strategy / Energy Management	Sets a 30-year roadmap, net-zero GHG goal by 2045, carbon-free grid goal by 2050, clean energy development, and electric vehicle expansion	Actions are support and pilot programs rather than strict mandates. It does not solve large data center energy demand (Loudoun Government).
Kaiserslautern	Masterplan 100% Klimaschutz	Guides the city toward becoming net zero, identifies energy reduction goals, solar potential, and 143 implementation measures	Success depends on long-term retrofits, resident participation, business cooperation, and renewable energy permitting (Stadt Kaiserslautern)
Freiburg	Climate Protection Policy + Klimafreundlich Wohnen	Targets climate neutrality by 2035, expands solar and wind, lowers energy demand, and funds energy-efficient building projects and renewable heating	Transferring the policy requires funding, public participation, and local renewable energy potential (Freiburg)
Portland	Climate Emergency Workplan + PCEF + Oregon clean-energy targets	Sets carbon-reduction targets, funds clean energy investments, and supports state clean electricity requirements	Implementation depends on political follow-through, oversight, program funding, and accountability systems (Portland Government)

2.4 Water Policies

The purpose of this category is to determine whether the city or region has taken steps to guarantee sustainable water use.

Table 4. Water Policies for Each City/Region

Place	Main Policies	What it does	Implementation Challenges
Loudoun County	Reclaimed water program + stormwater policy + data center water transparency	Reuses wastewater, sets stormwater controls, and public reporting for data center water use	Water management is split across agencies and remains mitigatory (Loudoun Water) (Kutner)
Kaiserslautern	Hochwasser- und Starkregen Vorsorgekonzept + stormwater utility standards	Focuses on flood sensitive areas, heavy rain preparedness, infiltration, reducing sealing, and keeping runoff close to pre-development conditions	Depends on redevelopment and enforcement (Stadt Kaiserslautern)
Freiburg	KLAK Water + GebäudeGrün hoch³ + split wastewater/stormwater fee + Vauban stormwater system	Uses rainwater retention, de-sealing, green roofs, green facades, stormwater fees, and decentralized infiltration	Transferring the policy requires local funding and redesigning streets (Freiburg)
Portland	2025 Stormwater Management Manual + One Water Initiative	Uses green stormwater infrastructure, sewer overflow reduction, and watershed protections	Successful transfer of the policy depends on funding and political support (Portland Government)

3. Methodology

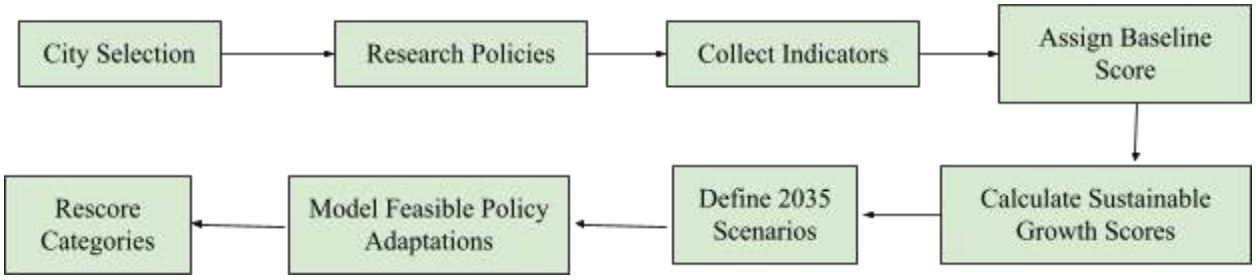


Figure 1. Flow Chart of Procedure.

3.1 Research Design and Case Selection

Loudoun County and Kaiserslautern were selected as the two cases as they represent different types of growth. Loudoun County is a car-dependent suburb in the United States facing a high growth pressure due to rapid data center expansion. Kaiserslautern is a compact German city with fairly strong land-use planning. Freiburg and Portland are ideal sustainable cities and thus were selected as policy benchmarks for the model.

3.2 Data Collection and Indicator Selection

Data was collected for five categories: transportation, land use, energy, water, and governance. Under transportation, quantitative indicators such as car commute share, public transit share, and mean travel time to work were determined for Loudoun County and Kaiserslautern. Under land use, quantitative indicators such as county land area, tree canopy

coverage, and impervious surface percentage were found. Under energy, quantitative indicators such as fossil fuel share, renewable energy share, and grid emissions intensity were determined. Under water, quantitative indicators such as drinking water demand and reclaimed water delivered were found. For governance, the policies for each city were researched and categorized.

3.3 Baseline Scoring Procedure

Based on the data, each city/region was assigned a baseline score from 1 to 5 in each of the five categories. A score of 1 or below represented weak sustainability performance and high long-term risk. A score of 5 represented best sustainability performance. The scoring process used both quantitative data and qualitative policy analysis.

Table 5. Rubric for Scoring Each Category

Category	Score 1	Score 2	Score 3	Score 4	Score 5
Transportation	Very car-dependent, limited public transit access	Mostly car-dependent, some alternatives exist	Mixed mobility system, moderate car use	Strong public transit, biking, and walking system	Benchmark-level low-car mobility, transit/bike/walk dominate trips
Land Use	High development pressure, high sprawl, weak green space protection	Some green space, but high development pressure	Meaningful green space but active growth pressure	Strong green-space protection or compact-growth policy	Benchmark-level compact development and green space as well as infrastructure
Energy	High fossil fuel dependence, rapid growth, major infrastructure pressure	Significant energy pressure, but some clean energy planning	Mixed energy profile with moderate renewable energy	Strong renewable and carbon-free progress or lower grid pressure	Benchmark-level clean-energy strategy with strong implementation
Water	High demand growth and limited mitigation or transparency	Rising demand with some planning but still pressure	Moderate pressure with some resilience planning	Strong water planning with adaptation, reuse, and supply resilience	Benchmark-level water efficiency, reuse, adaptation, and long-term resilience
Governance	No clear sustainability plan or weak funding and implementation	Plan exists but limited accountability or tracking	Formal plans with some implementation tools	Strong plans, policy tools, and public reporting	Strong plans, measurable targets, public progress tracking, funding, and accountability

3.4 Category Weights and Sustainable Growth Score (SGS)

$$SGS = 0.25T + 0.25L + 0.20E + 0.15W + 0.15G$$

where T = transportation, L = land use, E = energy, W = water, G = governance

After baseline scores were determined for each category, a weighted Sustainable Growth Score was calculated for each region. The formula is shown above, where transportation and land use were each weighed at 25 percent, energy was weighted at 20 percent, and water and

governance were each weighted at 15 percent. Transportation and land use were weighted the most as they shape infrastructure demand. Energy was weighted at 20 percent because of its impact on emissions and grid stress. Water and governance were weighted the least because success greatly depends on enforcement.

3.5 Scenario Definitions

Four scenarios were developed to represent the change to the weighted sustainability score under different policy conditions for Loudoun County and Kaiserslautern. Freiburg and Portland were not rescored because they served as benchmarks. The Current scenario represents the baseline condition for each implementation case. The Business as Usual 2035 scenario represents how sustainability could change if the policies of each city/region do not change. The Moderate Reform 2035 scenario estimates the effect of stronger implementation of existing policies. The Policy Transfer 2035 scenario models how performance could improve if each place adopted and adapted strategies from benchmark cities (as well as other cities in the study).

3.6 Policy Transfer Modeling

To model the Policy Transfer 2035 scenario, policy-transfer assumptions were developed within the five categories for both Loudoun County and Kaiserslautern. Often, a combination of policies were suggested for each category to maximize sustainability and feasibility. After the policy assumptions for each scenario were determined, category scores were estimated to reflect the impact of those changes. New weighted scores were then calculated for both Loudoun County and Kaiserslautern.

4. Results

4.1 Baseline Results

To make the scoring process transparent, all raw indicators were recorded in a separate table with each indicator's value, unit, year, source, source type, and notes. Based on these indicators and the policy research, each city/region was scored from 1 to 5 across five categories.

Table 6. Transportation Scores for Each City/Region

City/Region	Key indicators	Score	Justification
Loudoun County	60% car commute; 3.6% public transit/bike/walk share; 31.4 min mean commute (Loudoun Government)	**2.0**	Loudoun is very car-dependent and its public transit/bike/walk share is extremely low. It has some transit access through regional systems.
Kaiserslautern	40% car share; 60% public transit/bike/walk share; 26 min travel time (Stadt Kaiserslautern)	**4.0**	Kaiserslautern has low-car mobility planning and a fairly large portion of movement is by transit, biking, or walking.
Freiburg	21% car mode share; 79% public transit/bike/walk share; 34% bike share (ICLEI) (Kotas)	**4.5**	Freiburg is the transportation benchmark. Car use is low, biking is very high, and most trips happen through walking, biking, or public transit.

Portland	52% car commute; 37% public transit/bike/walk share; 7.8% bike share (Andersen) (Portland Bureau of Transportation) (Portland Government)	**3.5**	Portland has a stronger bike/transit/walk infrastructure than Loudoun and a higher low-car share as well. It has plans to lower car dependence.

Table 7. Land Use Scores for Each City/Region

City/Region	Key indicators	Score	Justification
Loudoun County	45% tree canopy; 11% impervious surface; 44% permeable non-canopy surface; 25.8% conservation easement share; high growth pressure (Loudoun Government) (Save Rural Loudoun)	**3.0**	Loudoun has a high tree canopy percentage and conservation easements. However, it faces significant growth pressure due to rapid data center expansion.
Kaiserslautern	44% tree/forest share; 30% impervious/settlement-transport land; 70% vegetation/water; 65.2% protected-area share; moderate growth pressure (Stadt Kaiserslautern)	**4.0**	Kaiserslautern has much green space and a lower growth pressure than Loudoun. There is settlement expansion.
Freiburg	153.04 km² land area; 32% impervious/urban development land; 42% forest/tree canopy; 27% permeable non-canopy surface; Vauban (Green City Freiburg)	**4.5**	Freiburg is very strong because of their long-term compact planning, green space preservation, and sustainable district models like Vauban.
Portland	345.6 km² land area; 40% impervious/urban development land; 29.8% tree canopy; 60% permeable surface; urban growth boundary (Portland Government) (Public Works)	**4.0**	Portland's tree canopy is lower than Loudoun's, but it has stronger long-term planning aimed at preventing urban sprawl.

Table 8. Energy Scores for Each City/Region

City/Region	Key indicators	Score	Justification
Loudoun County	61% fossil fuel share; 11% renewable share; 39% carbon-free share; 631 lbs CO_2/MWh; 3.1% annual electricity consumption increase; 40% electricity cost increase (Brown and Kemp) (EIA) (Loudoun Government)	**2.0**	Loudoun has high energy pressure because of the large-load electricity demand from data centers and heavy fossil fuel reliance.
Kaiserslautern	37.3% fossil fuel share; 62.7% renewable share; 62.7% carbon-free share; 334.5 gCO_2eq/kWh; 0.5% electricity consumption increase (Electricity Maps) (Stadt Kaiserslautern)	**3.5**	Germany's energy mix has a higher renewable share. However, some values are national proxies and aren't specific to Kaiserslautern.
Freiburg	Climate neutrality by 2038; renewable energy strategy; solar/wind/water/biomass focus; €120 million Future Fund (Freiburg)	**4.5**	Freiburg is a strong benchmark because it has clear goals, renewable energy planning, and allocated funding.
Portland	Renewable-energy targets; Oregon grid proxy; climate planning (Portland Government)	**4.0**	Portland has strong goals and a cleaner grid than Loudoun. However, some of its strength comes from targets rather than actual success.

Table 9. Water Scores for Each City/Region

City/Region	Key indicators	Score	Justification
Loudoun County	50 MGD current drinking water demand; 90 MGD projected 2040 demand; 2.1 MGD reclaimed water delivered (Kutner) (Loudoun Water)	**2.5**	Loudoun has some reclaimed water use, but the total water usage is expected to increase dramatically from 50 MGD to 90 MGD.
Kaiserslautern	3.16 MGD estimated current drinking water demand; 2040 groundwater/security planning; reclaimed water not found (Stadt Kaiserslautern)	**3.5**	Kaiserslautern has lower water demand and groundwater planning. There isn't much other data.
Freiburg	Green roofs and stormwater design in Vauban; sustainable stormwater management; near-natural watercourse development; soil protection (Freiburg)	**4.0**	Freiburg has strong water and green infrastructure planning. There isn't much other data.
Portland	Large-scale water planning; supply capacity; water infrastructure planning (Portland Government)	**4.0**	Portland is strong because it has good water-supply planning and infrastructure capacity.

Table 10. Governance Scores for Each City/Region

City/Region	Key indicators	Score	Justification
Loudoun County	2023 Energy Strategy; Phase 1 data center regulatory accountability; natural resources planning in development (Loudoun Government)	**3.5**	Loudoun has some plans and regulatory tools, but many policies are still developing.
Kaiserslautern	KLIMA 2050 plan; 2019 Climate Adaptation Plan (Stadt Kaiserslautern)	**4.0**	Kaiserslautern has good climate planning. Its main opportunity for improvement is implementation and tracking.
Freiburg	Climate neutrality by 2038; climate protection strategy; Future Fund; Vauban citizen participation (Freiburg)	**5.0**	Freiburg combines climate targets, implementation funding, sustainable district planning, and public participation.
Portland	Climate Emergency Workplan; climate progress reporting; renewable-energy commitments; public accountability mechanisms (Portland Government)	**4.5**	Portland has strong climate governance and public reporting.

Then, the weighted sustainable growth score was calculated for each city/region following the formula below.

Weighted Sustainable Growth Score = 0.25T + 0.25L + 0.20E + 0.15W + 0.15G

where T = transportation, L = land use, E = energy, W = water, G = governance

Loudoun Weighted Score = 0.25 (2.0) + 0.25 (3.0) + 0.20 (2.0) + 0.15 (2.5) + 0.15 (3.5) = 2.55

Table 11. Weighted Category and Summary Scores for Each City/Region

City/Region	Transportation	Land Use	Energy	Water	Governance	Weighted Score
Loudoun County	2.0	3.0	2.0	2.5	3.5	**2.55**
Kaiserslautern	4.0	4.0	3.5	3.5	4.0	**3.82**
Freiburg	4.5	4.5	4.5	4.0	5.0	**4.50**
Portland	3.5	4.0	4.0	4.0	4.5	**3.95**

A graph was then created to visualize these weighted scores for each city/region.

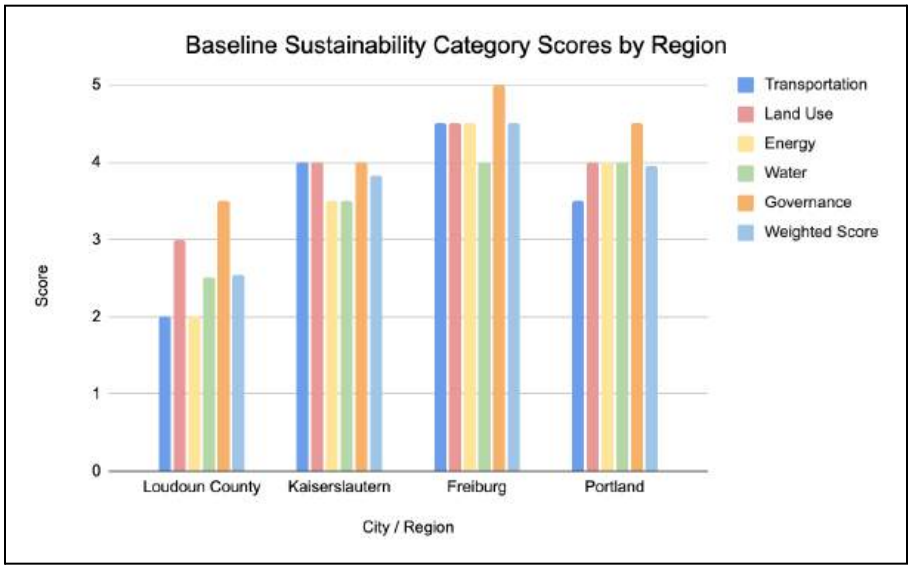


Figure 2. Baseline Sustainability Category Scores. Freiburg has the highest overall baseline scores and Loudoun has the lowest overall baseline scores.

4.2 Policy Transfer Model

Policy transfer tables determine what policies Loudoun County and Kaiserslautern could benefit from. They also indicate what a feasible 2035 adaptation could look like based on local conditions as well as the score impact of the change. The model does not assume that policies can be copied directly. Rather, it takes parts of policies from the benchmark cities as well as other cities in the study.

Table 12. Loudoun Policy Transfer Table for Each City/Region

Category	Current issue	Existing Policy	Benchmark Policies	Feasible 2035 adaptation	Score Impact

Transportation	Very high car dependence	2019 Countywide Transportation Plan	Kaiserslautern's Mobilitätsplan Klima+ 2030; Freiburg's low-car mobility model	1. Create low-car districts around Silver Line stations, town centers, schools, libraries, mixed-use areas. 2. Add protected bike lanes and safer pedestrian crossings in high-density corridors. 3. Include transportation demand management plans in large mixed-use development. 4. Improve shuttle access between popular sites. 5. Start with corridors instead of the entire county.	**2.0 → 3.0 or 3.25**
Land Use	High growth pressure from data centers and suburban development	Zoning Ordinance, Phase 1 Data Center Regulation	Portland's growth management; Freiburg's compact district planning	1. Require major developments (data centers) to show energy, water, transportation, and land-use impacts prior to approval. 2. Prioritize compact, mixed-use development. 3. Strengthen data center siting rules near community areas and water bodies. 4. Expand conservation in planned growth areas.	**3.0 → 3.75 or 4.0**
Energy	Rapid electricity demand	2023 Loudoun County Energy Strategy	Freiburg's climate protection funding; Portland's clean energy workplan	1. Require data centers and other major energy users to submit energy impact statements. 2. Require clean-energy plans for new large-load projects. Increase investment in clean energy projects, specifically through funding from large-load users/companies. 3. Require grid and peak load management plans. 4. Encourage onsite solar, battery, waste heat reuse, and energy efficient building design. 5. Report large-load energy demand trends.	**2.0 → 2.75 or 3.0**
Water	Water demand projected to rise dramatically	Reclaimed Water program / SB 553 for Virginia Data Center Water Transperancy	Kaiserslautern/Freiburg stormwater adaptation; Freiburg's water-sensitive design	1. Require large developments to submit water demand estimates before approval. 2. Require data centers to evaluate and implement low water cooling. 3. Increase reclaimed water usage and retention. 4. Expand stormwater retention infrastructure. 5. Add green infrastructure in more places (rain gardens, tree canopy, infiltration features).	**2.5 → 3.25 or 3.5**
Governance	Enforcement is uneven	Energy Strategy, Phase 1 data center regulation	Portland's public climate progress reporting; Freiburg's implementation of eco-friendly programs	1. Create a dashboard tracking data center approvals, square footage, energy demand, water use, emissions progress, and land-use change. 2. Track whether Energy Strategy actions are completed or delayed.	**3.5 → 4.25 or 4.5**

Table 13. Kaiserslautern Policy Transfer Table for Each City/Region

Category	Current Issue	Existing Policy	Benchmark Policies	Feasible 2035 adaptation	Score Impact

Transportation	Uneven funding; disjointed projects	Mobilitätsplan Klima 2030	Portland public reporting model; Freiburg cycling implementation	1. Close the gaps identified in the mobility plan on bike routes. 2. Build more protected bike lanes where street width allows. 3. Open more one-way streets to cyclists in the opposite direction. 4. Improve bus frequency and access to the university area, PRE-Park, the main station, and job centers. 5. Add bike parking, shared bikes, EV charging, and pedestrian safety features near common transit stops. 6. Require yearly transportation estimates for new housing, industrial, and commercial projects.	**4.0 → 4.5**
Land Use	Some urban sprawl and lack of public tracking	Flächennutzungsplan 2025	Portland urban growth management; Freiburg compact districts: Dietenbach / Vauban	1. Track new housing, brownfield reuse, green space, forest, protected green areas, and transport land added each year. 2. Require ecological compensation when new land is sealed. 3. Add mixed-use requirements in growth areas so housing, services, schools, and jobs are closer together. 4. Use concept-based land allocation for major public land projects, so land is allocated partly based on sustainability. 5. Set a target to reduce net new land sealing by 2035.	**4.0 → 4.4**
Energy	Limited public input, cooperation, and renewable energy investment	Masterplan 100% Klimaschutz	Freiburg climate funding/implementation model; Portland climate progress reporting	1. Track progress toward Masterplan targets. Prioritize A-measures. 2. Create monetary incentives for residential and commercial building retrofits. 3. Prioritize heat-pump adoption and renewable heating when upgrading buildings. 4. Exemplify public buildings to influence resident adoption. 5. Implement solar and storage systems in public facilities.	**3.5 → 4.2**
Water	Uneven de-sealing and new demand	Hochwasser- und Starkregenvorsorgekonzept	Freiburg KLAK water planning; Freiburg green-blue infrastructure; Vauban decentralized stormwater design	1. Track heavy-rain risk zones and completed flood-prevention projects. 2. Expand de-sealing programs in flood risk areas. 3. Require rainwater retention and infiltration during redevelopment. 4. Add green roofs and facades on public buildings and incentivize them on private buildings. 5. Convert selected paved areas into rain gardens, infiltration trenches, and tree pits to lower runoff. 6. Implement a decentralized stormwater design in new districts.	**3.5 → 4.2**
Governance	Not measurable or coordinated well	Mobilitätsplan Klima+ 2030; Flächennutzungsplan 2025; Masterplan 100% Klimaschutz	Portland Climate Emergency Workplan; Loudoun public infrastructure transparency; Freiburg climate funding	1. Publish an annual public report addressing transportation, land use, energy, and water. 2. Publish a dashboard showing completed, delayed, and voided actions. 3. Connect climate goals to yearly implementation targets. 4. Tie budgeting to the highest-priority actions. 5. Require every major development to show how it bolsters climate goals.	**4.0 → 4.7**

4.3 Scenario Results

After the baseline scores were calculated, the model was applied to four scenarios that explore how adopting these policies to varying extents would impact the weighted sustainability score for both Loudoun County and Kaiserslautern. The Current scenario represents the baseline score. The Business as Usual 2035 scenario estimates what could happen if current growth pressures continue with minimal policy changes. The Moderate Reform 2035 scenario estimates what could happen if areas strengthen implementation of current policies. The Policy Transfer 2035 scenario estimates changes to the sustainability score based on feasible adaptation and implementation of the selected policies indicated in Tables 12 and 13. The weighted score was calculated using the same formula as before.

Table 14. Loudoun Scenario Scores

Category	Current	Business as Usual 2035	Moderate Reform 2035	Policy Transfer 2035
Transportation	2.0	2.0	2.625	3.25
Land Use	3.0	2.25	3	3.75
Energy	2.0	1.5	2.25	3.0
Water	2.5	2.0	2.75	3.5
Governance	3.5	3.2	3.85	4.5
Weighted Score	2.55	2.14	2.85	3.55

Table 15. Kaiserslautern Scenario Scores

Category	Current	Business as Usual 2035	Moderate Reform 2035	Policy Transfer 2035
Transportation	4.0	4.0	4.25	4.5
Land Use	4.0	3.8	4.1	4.4
Energy	3.5	3.6	3.9	4.2
Water	3.5	3.5	3.85	4.2
Governance	4.0	4.0	4.35	4.7
Weighted Score	3.83	3.80	4.10	4.40

A graph was then created to visualize these weighted scores for each city/region.

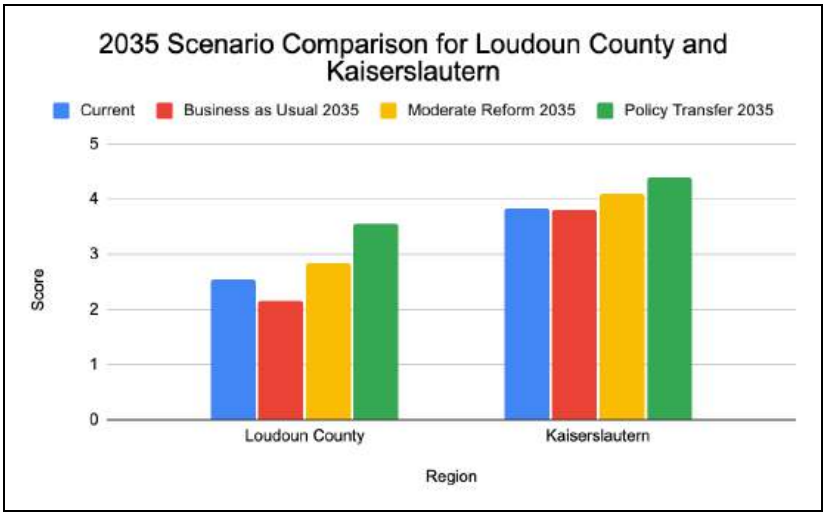


Figure 3. 2035 Scenario Comparison Visualization. The Policy Transfer Model produces the highest overall sustainable growth scores for both Kaiserslautern and Loudoun.

Radar charts were also created to indicate which sustainability category each place performed the best and worst for under the Current and Policy Transfer 2035 models.

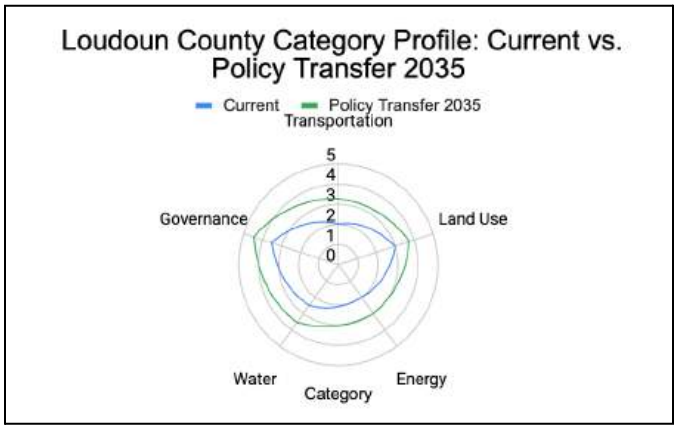


Figure 4. Loudoun County Radar Chart. Under the Policy Transfer scenario, Loudoun improves sustainability performance in all categories.

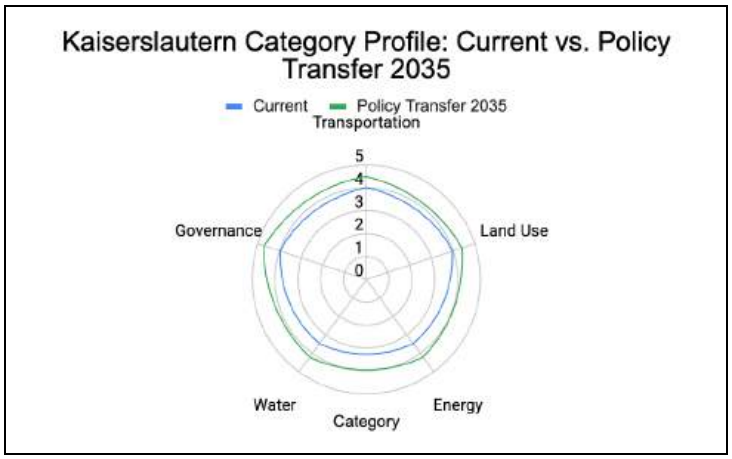


Figure 5. Kaiserslautern Radar Chart. Under the Policy Transfer scenario, Kaiserslautern improves in sustainability performance across all categories although the improvement is less than Loudoun's.

A sensitivity analysis was performed on the data to determine whether the results depended too much on the chosen weights in the Sustainability Growth Score. The Policy Transfer 2035 model was recalculated under three other weighting systems: equal weights, an energy-water stress model, and a governance-heavy model, as represented in Tables 16 and 17.

Table 16. Sensitivity Analysis Weighting Systems

Weight Test	Transportation	Land Use	Energy	Water	Governance
Original model	25%	25%	20%	15%	15%
Equal weights	20%	20%	20%	20%	20%
Energy-Water	20%	20%	25%	25%	10%
Governance	20%	20%	15%	15%	30%

Table 17. Sensitivity Analysis Results

Weight Test	Loudoun Policy Transfer 2035	Kaiserslautern Policy Transfer 2035
Original model	3.55	4.40
Equal weights	3.60	4.40
Energy-Water Stress	3.50	4.31
Governance	3.73	4.49

Discussion

The purpose of this project was to examine how Loudoun County and Kaiserslautern can improve sustainable growth outcomes by 2035 by adapting policies from U.S. and German benchmark cities. The results suggest that the project was successful as policy transfer models were developed for both Loudoun and Kaiserslautern. The adapted policies improved both Sustainable Growth Scores. Loudoun County increased from 2.55 under current conditions to 3.55. Kaiserslautern increased from 3.83 to 4.40. Sustainability policies were found to be more effective when they are adapted to local conditions. Loudoun County and Kaiserslautern have different growth pressures, and both regions are different from the benchmark cities of Freiburg and Portland. Thus, successful policy elements from each sustainability category were selected, adapted, and combined to develop 2035 sustainable growth score expectations for both Loudoun County and Kaiserslautern. This approach helped create the Policy Transfer model for each region and craft feasible policies that could actually improve sustainability.

Loudoun was found to perform weakest in the transportation and energy sustainability categories, as shown in Table 11. This is representative of the county's high car dependence and rapid data center expansion. The Policy Transfer Model in Table 12 indicates Loudoun's most feasible improvements that would improve its sustainability. These improvements all come from policies that are in place in Freiburg, Portland, or Kaiserslautern. For transportation, Loudoun should implement low-car districts near Silver Line stations, town centers, schools, and mixed-use areas, modeled after Freiburg. For land use, Loudoun should require data centers to submit energy, water transportation, and land-use impact reviews prior to approval. For energy, Loudoun should issue a mandate requiring large-load users to submit clean-energy plans, incentivizing funding for renewable energy and battery storage projects. For water, Loudoun should expand reclaimed water use, require low-water cooling in data centers, increase stormwater retention, and add rain gardens on buildings like data centers.

Although Kaiserslautern performed better than Loudoun overall, Kaiserslautern still performed the weakest in the energy and water categories, as shown in Table 11. The Policy Transfer Model in Table 13 displays Kaiserslautern's most feasible improvements that bolster its sustainability score. For transportation, Kaiserslautern should focus on closing bike and public transit network gaps by adding protected bike lanes and improving bus access to main transit areas. Kaiserslautern should also work on reconfiguring the bus route to minimize times and maximize convenience, incentivizing riders. For land use, Kaiserslautern should track brownfield use, require ecological compensation when land is sealed, add mixed-use requirements in growth areas, and set a target to reduce new land sealing by 2035. For energy, Kaiserslautern should prioritize funding building retrofits and adding solar, storage, and heat-pumps to public facilities (especially during upgrades). For water, Kaiserslautern should expand de-sealing and require rainwater retention. They should also continue to add green roofs and facades.

The radar charts further the study's findings by visually displaying the changes between the Current and Policy Transfer scenarios. Loudoun's radar chart in Figure 4 especially expands outward in the transportation, energy, and water categories, connecting to the policies proposed

in the Policy Transfer Model. Kaiserslautern's radar chart in Figure 5 also expands under Policy Transfer, but the change is more gradual.

However, the model had several limitations. First, the category weights used to calculate the Sustainable Growth Score were selected based on the relative importance of each category. The sensitivity analysis in Table 17 addresses this limitation. The model was recalculated using equal weights, an energy-water stress model, and a governance-heavy model. Across all three tests, the Policy Transfer scenario produced the highest modeled scores. This supports the model's conclusion. Another limitation was that the scoring system combined quantitative data with qualitative observations. Loudoun County, Kaiserslautern, Freiburg, and Portland do not all have the same amount of data, making some indicators easier to compare than others. In fact, some data relied on proxies, such as Kaiserslautern's energy mix. Finally, the model does not forecast exactly what will happen in 2035. Instead, it estimates how policies could impact sustainability outcomes if implemented. Success depends on funding, political conditions, public acceptance, coordination, and implementation.

Overall, the Policy Transfer models provide catered policy suggestions for both Loudoun County and Kaiserslautern that could improve their sustainable growth outcomes by 2035. Loudoun's reforms focus on managing infrastructure-heavy growth, while Kaiserslautern's reforms focus on strengthening implementation and filling the gaps in current policies.

Works Cited

Andersen, Michael. “Bike commute numbers ebb nationwide; in Portland, they’re flat.” 2018, https://bikeportland.org/2018/09/25/bike-commute-numbers-ebb-nationwide-in-portland-theyre-flat-290026.

Aquabench. “PREIS- UND TARIFINFORMATIONSBLATT WASSER – WIRTSCHAFTSJAHR 2025.” 2024, https://mkuem.rlp.de/fileadmin/14/Themen/Wasser/Benchmarkinginitiative_2025/Benchmarkinginitiative_Wasser/PTiB_WV_2025_SWK_Stadtwerke_KaiserslauternVersorgungs-AG.pdf.

Brown, Neel, and John Kemp. “The Virginia Challenge: Meeting Energy Demand Affordably.” Progressive Policy Institute, 2026, https://www.progressivepolicy.org/wp-content/uploads/2026/01/PPI_The-Virginia-Challenge.pdf.

Census Reporter. “Loudoun County, VA.” 2024, https://censusreporter.org/profiles/05000US51107-loudoun-county-va/.

City Population EU. “Kaiserslautern.” 2025, https://www.citypopulation.de/en/germany/rheinlandpfalz/kaiserslautern_stadt/07312000__kaiserslautern/.

EIA. “Virginia Electricity Profile 2024.” 2024, https://www.eia.gov/electricity/state/virginia/.

Electricity Maps. “Electricity Grid Review 2025: Germany.” 2025, https://www.electricitymaps.com/grid-in-review-2025/germany

Europawahl 2024. “Kaiserslautern, Stadt.” 2024, https://www.bundeswahlleiterin.de/europawahlen/2024/strukturdaten/bund-99/land-7/kreis-7312.html.

European Parliament election 2024. “Freiburg im Breisgau, Stadtkreis.” 2023, https://www.bundeswahlleiterin.de/en/europawahlen/2024/strukturdaten/bund-99/land-8/kreis-8311.html.

Faguy, Ana. “A humming annoyance or jobs boom? Life next to 199 data centres.” 2025, https://www.bbc.com/news/articles/c93dnnxewdvo.

Freiburg. “Energie und Klimaschutz.” 2025, https://www.freiburg.de/pb/232049.html.

Freiburg. “Förderprogramm Klimafreundlich Wohnen.” https://www.freiburg.de/pb/232441.html.

Freiburg. “Handlungsfeld Regenwasser.” 2024, https://www.freiburg.de/pb/2174138.html.

Green City Freiburg. “Urban planning.” 2020, https://greencity.freiburg.de/pb/,Len/1652540.html.

ICLEI. “Freiburg im Breisgau, Deutschland." 2020, https://sustainablemobility.iclei.org/ecomobility-alliance/freiburg-germany/.

Kaiserslautern. “Bauleitplanungen, Bauvorhaben, Flächennutzung.” 2024, https://www.kaiserslautern.de/mb/themen/umwelt/natur/pdf/3_8_bauleitplanung_und_bauvorhaben.pdf.

Kaiserslautern. “Schutzgebiete und Objekte.” 2025, https://www.kaiserslautern.de/sozial_leben_wohnen/umwelt/natur_und_landschaft/schutzgebiete/index.html.de.

Koch, Florian, and Kerstin Krellenberg. “How to Contextualize SDG 11? Looking at Indicators for Sustainable Urban Development in Germany.” International Journal of Geo-Information, vol. 7, no. 12, 2018, https://doi.org/10.3390/ijgi7120464.

Kotas, Dominic. “Freiburg’s Cycling Concept 2020.” ICLEI, 2015, https://talkofthecities.iclei.org/freiburgs-cycling-concept-2020/.
Kutner, Brad. “New Virginia law requires data center water usage be made public.” 2026, https://www.wmra.org/2026-06-11/new-virginia-law-requires-data-center-water-usage-be-made-public.
Loudoun Government. “Loudoun County 2019 Countywide Transportation Plan.” 20 June 2019, https://www.loudoun.gov/DocumentCenter/View/152287/CTP---Combined-with-small-maps-bookmarked.
Loudoun Government. “Loudoun County Energy Strategy.” 2023, https://www.loudoun.gov/DocumentCenter/View/174600/Loudoun-County-Energy-Strategy-Final-Report---2023.
Loudoun Government. “Loudoun County Zoning Ordinance.” 13 December 2023, https://online.encodeplus.com/regs/loudouncounty-va-crosswalk/page/homepage.
Loudoun Government. “Tree Canopy.” 2025, https://www.loudoun.gov/TreeCanopy.
Loudoun Water. “Potomac Water Supply Program.” 2026, https://www.loudounwater.org/potomac-water-supply-program.
Loudoun Water. “Reclaimed Water Program.” 2010, https://www.loudounwater.org/commercial-customers/reclaimed-water-program.
Macro Trends. “Portland Metro Area Population | Historical Data | Chart | 1950-2026.” 2026, https://www.macrotrends.net/datasets/global-metrics/cities/23102/portland/population.
McDarris, Alisha. “This may be one of Germany’s most sustainable cities. So what is Freiburg doing right?” 2024, https://adventure.com/freiburg-germany-blueprint-sustainable-city/.
Michalina, Denis, et al. “Sustainable Urban Development: A Review of Urban Sustainability Indicator Frameworks.” Sustainability, vol. 13, no. 16, 2021, https://doi.org/10.3390/su13169348.
Portland Bureau of Transportation. “2023 Portland Bicycle Counts.” 13 February 2024, https://www.portland.gov/transportation/bicycle-committee/documents/presentation-preview-2023-bicycle-count-report/download.
Portland Government. “Climate Emergency Workplan.” 2022, https://www.portland.gov/bps/climate-action/climate-emergency/climate-emergency-workplan.
Portland Government. “Climate Investment Plan: PCEF's five-year funding road map to climate action.” December 2024, https://www.portland.gov/bps/cleanenergy/climate-investment-plan.
Portland Government. “One Water: Integrating Water, Sewer, and Stormwater Services.” 2024, https://www.portland.gov/water/one-water.
Portland Government. “Plans, Reports, Maps, and Guidebooks.” 2025, https://www.portland.gov/trees/plansandreports.
Portland Government. “Title 33 Planning and Zoning.” December 2021, https://www.portland.gov/code/33.
Portland Government. “Transportation System Plan (TSP).” 2018, https://www.portland.gov/transportation/planning/tsp.

Portland Government. “2025 Stormwater Management Manual.” 2025, https://www.portland.gov/bes/stormwater/2025-stormwater-management-manual.
“Portland, Oregon.” 2012, https://www.nrdc.org/sites/default/files/RooftopstoRivers_Portland.pdf.
Public Works. “Commute Methods: Percentage of Portland commuters who drove alone.” 2024, https://embed.clearimpact.com/Measure/Embed/100717780.
Save Rural Loudoun. “Unconstrained Growth in the Rural Policy Area.” 2023, https://www.saveruralloudoun.org/the-issues.
Sense. “The 15 Most Sustainable Cities In The US.” Sense, 2025, https://sense.com/consumer-blog/the-15-most-sustainable-cities-in-the-us/.
Stadt Kaiserslautern. “Flächennutzungsplan 2025 inklusive Teiländerungen.” 2017, https://www.kaiserslautern.de/insek/dokumentation/vorliegendeplanungen/061567/index.html.de.
Stadt Kaiserslautern. “Hochwasser und ein starkes Vorsorgekonzept." 2026, https://www.kaiserslautern.de/sozial_leben_wohnen/umwelt/wasser/hochwasser/hochwasser_starkregenvorsorgekonzept/index.html.de.
Stadt Kaiserslautern. “Masterplan 100 % Klimaschutz: Energiewende Kaiserslautern Gemeinsam zum Ziel.” 2017, https://www.kaiserslautern.de/mb/themen/umwelt/klima/20171116_masterplan_gesamtdokument.pdf.
Stadt Kaiserslautern. “Mobilitätsplan Klima 2030.” 16 August 2018, https://www.kaiserslautern.de/mb/themen/verkehr/mobilitaetsplan-neu/2018-08-16_kl_mobilitaetsplan-klima%2B2030_bericht.pdf.
United Nations Department of Economic and Social Affairs. “68% of the world population is projected to live in urban areas by 2050, says UN.” 2018, https://www.un.org/uk/desa/68-world-population-projected-live-urban-areas-2050-says-un.
Urban Agency. “Dietenbach.” 2023, https://urban-agency.com/dietenbach.
US Government. “Census Portland City, Oregon.” 2024, https://www.census.gov/quickfacts/fact/table/portlandcityoregon/PST045224.
US Government. “US Census Loudoun County Virginia.” 2025, https://www.census.gov/quickfacts/fact/table/loudouncountyvirginia/PST045224.
World Population Review. “Freiburg im Breisgau.” 2026, https://worldpopulationreview.com/cities/germany/freiburg-im-breisgau.
WWF. “Freiburg green city.” 2012, https://wwf.panda.org/wwf_news/?204419/Freiburg-green-city.
Zentrum Klimaanpassung. “Stadt Kaiserslautern – Klimaanpassungskonzept.” 2026, https://zentrum-klimaanpassung.de/wissen-klimaanpassung/praxisbeispiele/stadt-kaiserslautern-klimaanpassungskonzept.
Zukunfts Region. “Our region in numbers.” 2023, https://www.zukunftsregion-westpfalz.de/en/our-region-numbers.